

PCTWORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

DEX-0230



DEX-0237

CB

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :

**C12N 15/12, 15/85, C07K 14/435, C07H
21/04****A1**

(11) International Publication Number:

WO 98/07854

(43) International Publication Date:

26 February 1998 (26.02.98)(21) International Application Number: **PCT/US97/14637**(22) International Filing Date: **20 August 1997 (20.08.97)**

(30) Priority Data:

08/700,013**20 August 1996 (20.08.96)****US**(71) Applicant (for all designated States except US): **ALLELIX
NEUROSCIENCE, INC. [US/US]; 6850 Goreway Drive,
Mississauga, Ontario L4V 1V7 (CA).**

(72) Inventors; and

(75) Inventors/Applicants (for US only): **ALBERT, Vivian, R.
[US/US]; 65 Union Street, No. 31, Montclair, NJ 07042
(US). KOWALSKI, Leslie, R., Z. [US/US]; 202 Glen Drive,
Cedar Knolls, NJ 07927 (US). BORDEN, Laurence, A.
[US/US]; Apartment 7F, 160 Overlook Avenue, Hacken-
sack, NJ 07601 (US). McKELVY, Jeffrey, F. [US/US];
Apartment 18B, 275 Central Park West, Hackensack, NJ
07601 (US).**(74) Agent: **BLOOM, Allen; Dechert Price & Rhoads, P.O. Box
5218, Princeton, NJ 08543-5218 (US).**(81) Designated States: **AL, AM, AT, AU, AZ, BA, BB, BG, BR,
BY, CA, CH, CN, CU, CZ, DE, DK, EE, ES, FI, GB, GE,
GH, HU, IL, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR,
LS, LT, LU, LV, MD, MG, MK, MN, MW, MX, NO, NZ,
PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR,
TT, UA, UG, US, UZ, VN, YU, ZW, ARIPO patent (GH,
KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ,
BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE,
CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL,
PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN,
ML, MR, NE, SN, TD, TG).****Published***With international search report.*(54) Title: **HUMAN GLYCINE TRANSPORTER**

(57) Abstract

Provided are nucleic acids and proteins derived from the sequences of the human GlyT-2 transporter of the amino acid glycine.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	ES	Spain	LS	Lesotho	SI	Slovenia
AM	Armenia	FI	Finland	LT	Lithuania	SK	Slovakia
AT	Austria	FR	France	LU	Luxembourg	SN	Senegal
AU	Australia	GA	Gabon	LV	Latvia	SZ	Swaziland
AZ	Azerbaijan	GB	United Kingdom	MC	Monaco	TD	Chad
BA	Bosnia and Herzegovina	GE	Georgia	MD	Republic of Moldova	TG	Togo
BB	Barbados	GH	Ghana	MG	Madagascar	TJ	Tajikistan
BE	Belgium	GN	Guinea	MK	The former Yugoslav Republic of Macedonia	TM	Turkmenistan
BF	Burkina Faso	GR	Greece			TR	Turkey
BG	Bulgaria	HU	Hungary	ML	Mali	TT	Trinidad and Tobago
BJ	Benin	IE	Ireland	MN	Mongolia	UA	Ukraine
BR	Brazil	IL	Israel	MR	Mauritania	UG	Uganda
BY	Belarus	IS	Iceland	MW	Malawi	US	United States of America
CA	Canada	IT	Italy	MX	Mexico	UZ	Uzbekistan
CF	Central African Republic	JP	Japan	NE	Niger	VN	Viet Nam
CG	Congo	KE	Kenya	NL	Netherlands	YU	Yugoslavia
CH	Switzerland	KG	Kyrgyzstan	NO	Norway	ZW	Zimbabwe
CI	Côte d'Ivoire	KP	Democratic People's Republic of Korea	NZ	New Zealand		
CM	Cameroon			PL	Poland		
CN	China	KR	Republic of Korea	PT	Portugal		
CU	Cuba	KZ	Kazakhstan	RO	Romania		
CZ	Czech Republic	LC	Saint Lucia	RU	Russian Federation		
DE	Germany	LI	Liechtenstein	SD	Sudan		
DK	Denmark	LK	Sri Lanka	SE	Sweden		
EE	Estonia	LR	Liberia	SG	Singapore		

- 1 -

HUMAN GLYCINE TRANSPORTER

This application is related to the following co-pending applications: "Glycine Transporter-Transfected Cells and Uses Thereof," Attorney Docket No. 317743-105, Serial No. 08/655,836, filed May 31, 1996; "Pharmaceutical For Treatment Of Neurological And Neuropsychiatric Disorders," Attorney Docket No. 317743-103, Serial
5 No. 08/656,063, filed May 31, 1996; "Pharmaceutical For Treatment of Neuropsychiatric Disorders," Attorney Docket No. 317743-106, Serial No. 08/655,912, filed May 31, 1996; and "Pharmaceutical For Treating Of Neurological and Neuropsychiatric Disorders," Attorney Docket No. 317743-107, Serial No. 08/655,847, filed May 31, 1996.

10 The present invention relates to nucleic acid encoding the "GlyT-2" member of the family of human glycine transporters, to the isolated protein encoded by the nucleic acid, and to the field of drug discovery.

Synaptic transmission is a complex form of intercellular communication that involves a considerable array of specialized structures in both the pre- and post-synaptic neuron. High-affinity neurotransmitter transporters are one such component, located on
15 the pre-synaptic terminal and surrounding glial cells (Kanner and Schuldiner, *CRC Critical Reviews in Biochemistry* 22: 1032, 1987). Transporters sequester neurotransmitter from the synapse, thereby regulating the concentration of neurotransmitter in the synapse, as well as its duration in the synapse, which together influence the magnitude of synaptic transmission. By preventing the spread of transmitter
20 to neighboring synapses, transporters maintain the fidelity of synaptic transmission. Further, by sequestering released transmitter into the presynaptic terminal, transporters allow for transmitter reutilization.

Neurotransmitter transport is dependent on extracellular sodium and the voltage difference across the membrane: under conditions of intense neuronal firing, as
25 for example during a seizure, transporters can function in reverse, releasing neurotransmitter in a calcium-independent non-exocytotic manner (Attwell et al., *Neuron* 11: 401-407, 1993). Pharmacologic modulation of neurotransmitter transporters thus provides a means for modifying synaptic activity, which provides useful therapy for the treatment of neurological and psychiatric disturbances.

30

- 2 -

The amino acid glycine is a major neurotransmitter in the mammalian nervous system, functioning at both inhibitory and excitatory synapses. By the phrase "nervous system," both the central and peripheral portions of the nervous system are intended. The distinct inhibitory and excitatory functions of glycine are mediated by two
5 different types of receptor, each of which is associated with a different class of glycine transporter. The inhibitory actions of glycine are mediated by glycine receptors that are sensitive to the convulsant alkaloid, strychnine, and are thus referred to as "strychnine-sensitive". Such receptors contain an intrinsic chloride channel that is opened upon
10 binding of glycine to the receptor; by increasing chloride conductance, the threshold for firing of an action potential is increased. Strychnine-sensitive glycine receptors are found predominantly in the spinal cord and brainstem, and pharmacological agents that enhance the activation of such receptors will thus increase inhibitory neurotransmission in these regions.

Glycine functions in excitatory transmission by modulating the actions of
15 glutamate, the major excitatory neurotransmitter in the central nervous system. See Johnson and Ascher, *Nature* 325: 529-531, 1987; Fletcher et al., *Glycine Transmission* Otterson and Storm-Mathisen, eds., 1990, pp. 193-219. Specifically, glycine is an obligatory co-agonist at the class of glutamate receptor termed N-methyl-D-aspartate (NMDA) receptor. Activation of NMDA receptors on a neuron increases sodium and
20 calcium conductance, which depolarizes the neuron, thereby increasing the likelihood that the neuron will fire an action potential. NMDA receptors are widely distributed throughout the brain, with a particularly high density in the cerebral cortex and hippocampal formation.

Molecular cloning has revealed the existence in mammalian brains of two
25 classes of glycine transporters, termed GlyT-1 and GlyT-2. GlyT-1 is found predominantly in the forebrain, and its distribution corresponds to that of glutamatergic pathways and NMDA receptors (Smith, et al., *Neuron* 8: 927-935, 1992). The distribution of GlyT-2 differs; this transporter is found predominantly in the brain stem and spinal cord, and its distribution corresponds closely to that of strychnine-sensitive
30 glycine receptors. Liu et al., *J. Biol. Chem.* 268: 22802-22808, 1993; Jursky and Nelson, *J. Neurochem.* 64: 1026-1033, 1995. These observations are consistent with the view that, by regulating the synaptic levels of glycine, GlyT-1 and GlyT-2 preferentially influence the activity of NMDA receptors and strychnine-sensitive glycine receptors, respectively.

- 3 -

Sequence comparisons of GlyT-1 and GlyT-2 have revealed that these glycine transporters are members of a broader family of sodium-dependent neurotransmitter transporters, including, for example, transporters specific for γ -amino-n-butyric acid (GABA) and others. Uhl, *Trends in Neuroscience* 15: 265-268, 1992; Clark and Amara, *BioEssays* 15: 323-332, 1993. Overall, each of these transporters includes 12 putative transmembrane domains that predominantly contain hydrophobic amino acids. Comparing rat GlyT-1 to rat GlyT-2, using the Lipman-Pearson FASTA algorithm, reveals a 51% amino acid sequence identity and a 55% nucleic acid sequence identity. Comparison of the sequence of human GlyT-1 with rat GlyT-2 reveals a 51% amino acid sequence identity and a 53-55% nucleic acid sequence identity, with the range of values for nucleic acid sequence identity resulting from the existence of three variant forms of GlyT-1.

Compounds that inhibit or activate glycine transporters would be expected to alter receptor function, and provide therapeutic benefits in a variety of disease states. For example, inhibition of GlyT-2 can be used to diminish the activity of neurons having strychnine-sensitive glycine receptors via increasing synaptic levels of glycine, thus diminishing the transmission of pain-related (*i.e.*, nociceptive) information in the spinal cord, which has been shown to be mediated by these receptors. Yaksh, *Pain* 111-123, 1989. Additionally, enhancing inhibitory glycinergic transmission through strychnine-sensitive glycine receptors in the spinal cord can be used to decrease muscle hyperactivity, which is useful in treating diseases or conditions associated with increased muscle contraction, such as spasticity, myoclonus (which relates to rapid muscle spasms), and epilepsy (Truong et al., *Movement Disorders* 3: 77-87, 1988; Becker, *FASEB J.* 4: 2767-2774, 1990). Spasticity that can be treated via modulation of glycine receptors is associated with epilepsy, stroke, head trauma, multiple sclerosis, spinal cord injury, dystonia, and other conditions of illness and injury of the nervous system.

Summary of the Invention

In a first embodiment, the invention provides a nucleic acid encoding a glycine transporter having at least about 96% sequence identity with the protein sequence of SEQ ID NO:27 or with a sequence corresponding to the protein sequence of SEQ ID NO:27 except that it has one or more of the following substitutions (1) Gly¹⁰² to Ser, (2) Ser¹²⁴ to Phe, (3) Ile²⁷⁹ to Asn, (4) Arg³⁹³ to Gly, (5) Lys⁴⁵⁷ to Asn, (6) Asp⁴⁶³ to Asn, (7) Cys⁶¹⁰ to Tyr, (8) Ile⁶¹¹ to Val, (9) Phe⁷³³ to Ser, (10) Ile⁷³⁵ to Val, (11) Phe⁷⁴⁵ to

- 4 -

Leu, (12) Val³⁰⁵ to Leu, (13) Thr³⁶⁶ to Ile or (14) Leu⁴⁰⁰ to Pro. Preferably, the sequence identity is at least about 97%, more preferably at least about 98%, yet more preferably at least about 99%, yet more preferably at least about 99.5%. In an embodiment of the invention, the sequence identity is 100%. Preferably, the encoded glycine transporter has no more than four amino acid differences in the region from amino acid 200 to 797 of reference protein sequence, where the reference sequence is SEQ ID NO:27 or of a sequence corresponding to the protein sequence of SEQ ID NO:27 except that it has one of the substitutions described above. More preferably, the encoded glycine transporter has no more than two such differences.

The invention also provides a vector comprising the nucleic acid described above. In one embodiment, the vector is effective to express a glycine transporter mRNA in at least one of a bacterial cell or a eukaryotic cell. In another embodiment of the invention, the vector is effective to express the mRNA in at least one of a yeast cell, a mammalian cell or an avian cell.

The invention further provides an isolated glycine transporter derived from transformed cells according to the invention, the transporter comprising the amino acid sequence encoded by the above-described nucleic acid or one to two contiguous portions of amino acid sequence encoded by such a nucleic acid, wherein the protein has glycine transporter activity and differs in sequence from the aligned segments of the rat transporter sequence. The phrase "contiguous sequence," as used herein, refers to uninterrupted portions of the relevant reference nucleic acid or amino acid sequence. Preferably, the glycine transporter protein of the present invention differs in sequence from the aligned segments of the rat transporter sequence by at least two amino acids, more preferably, at least four amino acids. Preferably, the contiguous sequences

comprise at least about 600 amino acids, more preferably at least about 700 amino acids, more preferably at least about 750 amino acids. In one embodiment, the transporter protein comprises all of the protein sequence encoded by the above-described nucleic acid. Preferably, the transporter protein comprises amino acid sequence set forth in the protein sequence of SEQ ID NO:27 or a sequence corresponding to the protein sequence of SEQ ID NO:27 except that it has one or more of the following substitutions (1) for Gly¹⁰², Ser, (2) for Ser¹²⁴, Phe, (3) for Ile²⁷⁹, Asn, (4) for Arg³⁹³, Gly, (5) for Lys⁴⁵⁷, Asn, (6) for Asp⁴⁶³, Asn, (7) for Cys⁶¹⁰, Tyr, (8) for Ile⁶¹¹, Val, (9) for Phe⁷³³, Ser, (10) for Ile⁷³⁵, Val, (11) for Phe⁷⁴⁵, Leu, (12) for Val³⁰⁵, Leu, (13) for Thr³⁶⁶, Ile or (14) for Leu⁴⁰⁰, Pro, or an amino acid sequence comprising one to two contiguous portions of

- 5 -

these sequences. In a preferred embodiment, the invention provides a glycine transporter and associated nucleic acids, vectors and methods, wherein the protein sequence comprises at least one of (1) Ser¹⁰², (2) Phe¹²⁴, (3) Asn²⁷⁹, (4) Gly³⁹³, (5) Asn⁴⁵⁷, (6) Asn⁴⁶³, (7) Tyr⁶¹⁰, (8) Val⁶¹¹, (9) Ser⁷³³, (10) Val⁷³⁵, (11) Leu²⁴⁵, (12) Leu³⁰⁵, (13) Ile³⁶⁶ and (14) Pro⁴⁰⁰. Preferably, the sequence comprises at least two of these amino acid residues, more preferably at least four, yet more preferably all of these amino acid residues:

In a second embodiment, the invention also provides a nucleic acid encoding a transporter protein having at least about 99.5% sequence identity with all or one to two contiguous portions of the amino acid sequence of SEQ ID NO:27 or with one to two continuous portions of an amino acid sequence corresponding to the protein sequence of SEQ ID NO:27 except that it has one or more of the following substitutions (1) Gly¹⁰² to Ser, (2) Ser¹²⁴ to Phe, (3) Ile²⁷⁹ to Asn, (4) Arg³⁹³ to Gly, (5) Lys⁴⁵⁷ to Asn, (6) Asp⁴⁶³ to Asn, (7) Cys⁶¹⁰ to Tyr, (8) Ile⁶¹¹ to Val, (9) Phe⁷³³ to Ser, (10) Ile⁷³⁵ to Val, (11) Phe²⁴⁵ to Leu, (12) Val³⁰⁵ to Leu, (13) Thr³⁶⁶ to Ile or (14) Leu⁴⁰⁰ to Pro, wherein the encoded protein has glycine transporter activity. Preferably, the contiguous sequences comprise at least about 600 amino acids, more preferably at least about 700 amino acids, more preferably at least about 750 amino acids. The invention also provides a vector comprising this nucleic acid. In one embodiment, the vector is effective to express a glycine transporter mRNA in at least one of a prokaryotic cell such as a bacterial cell or a eukaryotic cell. In another embodiment of the invention, the vector is effective to express the mRNA in at least one of a yeast cell, a mammalian cell or an avian cell.

The invention additionally provides a cell comprising a first extrinsically-derived nucleic acid according to the first embodiment or a second extrinsically-derived nucleic acid encoding a transporter protein having at least about 99.5% sequence identity with one to two contiguous portions of the protein sequence of SEQ ID NO:27 or of a sequence corresponding to the protein sequence of SEQ ID NO:27 except that it has one or more of the following substitutions (1) Gly¹⁰² to Ser, (2) Ser¹²⁴ to Phe, (3) Ile²⁷⁹ to Asn, (4) Arg³⁹³ to Gly, (5) Lys⁴⁵⁷ to Asn, (6) Asp⁴⁶³ to Asn, (7) Cys⁶¹⁰ to Tyr, (8) Ile⁶¹¹ to Val, (9) Phe⁷³³ to Ser, (10) Ile⁷³⁵ to Val, (11) Phe²⁴⁵ to Leu, (12) Val³⁰⁵ to Leu, (13) Thr³⁶⁶ to Ile or (14) Leu⁴⁰⁰ to Pro, wherein the encoded protein has glycine transporter activity. In one embodiment, the cell expresses a glycine transporter from the nucleic acid. Preferably, the nucleic acid is functionally associated with a promoter that is operative in the cell. In an embodiment of the invention, the promoter is an inducible

- 6 -

promoter.

The invention also provides a method of producing a glycine transporter comprising growing the cells described in the previous paragraph. This method can further comprise at least one of (a) isolating membranes from said cells, which
 5 membranes comprise the glycine transporter or (b) extracting a protein fraction from the cells, which fraction comprises the glycine transporter.

An embodiment of the invention provides a method for characterizing a bioactive agent for treatment of a nervous system disorder or condition or for identifying bioactive agents for treatment of a nervous system disorder or condition, the method
 10 comprising (a) providing a first assay composition comprising (i) a cell as described above or (ii) an isolated glycine transporter protein comprising the amino acid sequence encoded by the first or second extrinsically-derived nucleic acids described above, (b) contacting the first assay composition with the bioactive agent or a prospective bioactive agent, and measuring the amount of glycine transport exhibited by the assay composition.
 15 Preferably, the method further comprises comparing the amount of glycine transport exhibited by the first assay composition with the amount of glycine transport exhibited by a second such assay composition that is treated the same as the first assay composition except that it is not contacted with the bioactive agent or prospective bioactive agent. The method can be used for characterizing bioactive agents where the nervous system
 20 disorder or condition is one of the group consisting of (a) pain, (b) spasticity, (c) myoclonus, (d) muscle spasm, (e) muscle hyperactivity or (f) epilepsy. In a preferred embodiment, the spasticity for which the bioactive agent is characterized is associated with stroke, head trauma, neuronal cell death, multiple sclerosis, spinal cord injury, dystonia, Huntington's disease or amyotrophic lateral sclerosis.

25 The invention further provides a nucleic acid that hybridizes with a reference nucleic acid sequence which is SEQ ID NO:26 or a sequence that varies from the nucleic acid sequence of SEQ ID NO:26 by having one or more of the following substitutions (a) T⁶ to C, (b) G³⁰⁴ to A, (c) C³⁷¹ to T, (d) C⁵⁷¹ to T, (e) T⁸³⁶ to A, (f) A¹¹¹⁶ to G, (g) A¹¹⁷⁷ to G, (h) G¹³⁷¹ to C, (i) G¹³⁸⁷ to A, (j) G¹⁸²⁹ to A, (k) A¹⁸³¹ to G, (l) G²¹⁰³ to A,
 30 (m) T²¹⁹⁸ to C, (n) A²²⁰³ to G, (o) C³⁴² to G, (p) C³⁵² to T, (q) T⁷³³ to C, (r) A⁷⁷⁷ to G, (s) G⁹¹³ to C, (t) G⁹⁵¹ to A, (u) C¹⁰⁹⁷ to T or (v) T¹¹⁹⁹ to C, under conditions of sufficient stringency to exclude hybridizations with (a) the sequence for a rat or mouse GlyT-2 transporter or (b) the sequence for a mammalian GlyT-1 transporter. Preferably, the nucleic acid sequence is at least about 18 nucleotides in length and has at least about

- 7 -

95% sequence identity with a sequence embedded in the reference nucleic acid sequence. Preferably the nucleic acid sequence is at least about 40 nucleotides in length, more preferably at least about 100 nucleotides in length. Preferably the nucleic acid sequence has at least about 97% sequence identity with the above-recited reference sequence, more preferably 99% sequence identity. Preferably, the nucleic acid is a PCR primer and the stringent conditions are PCR conditions effective to amplify a human GlyT-2 sequence but not to amplify (a) the sequence for a rat or mouse GlyT-2 transporter or (b) the sequence for a mammalian GlyT-1 transporter.

Further, the invention provides a nucleic acid of at least about 18 nucleotides in length comprising a contiguous sequence from the coding or noncoding strand of a human GlyT-2 gene or cDNA, wherein the contiguous sequence has at least 1 sequence difference when compared with the rat GlyT-2 gene sequence that aligns with the contiguous sequence. Preferably the nucleic acid sequence is at least about 40 nucleotides in length, more preferably at least about 100 nucleotides in length. Preferably, the contiguous sequence has at least two differences, more preferably 3 differences when compared with the rat GlyT-2 gene sequence that aligns with the contiguous sequence.

Still further, the invention provides an antisense molecule comprising a contiguous sequence from a coding or non-coding strand of a human gene or cDNA for GlyT-2 which is effective when administered to a cell, tissue, organ or animal to reduce the expression of GlyT-2 in the cell or in a cell of the tissue, organ or animal, wherein the contiguous sequence has at least 1 sequence difference when compared with the rat GlyT-2 gene sequence that aligns with said contiguous sequence. Preferably, the contiguous sequence has at least two differences, more preferably 3 differences when compared with the rat GlyT-2 gene sequence that aligns with the contiguous sequence. The phrase "antisense molecule," is used herein to refer to a molecule designed to bind genomic DNA or mRNA to interfere in transcription or translation, including interfering with mRNA stability. Preferably, the contiguous sequence is at least about 15 nucleotides in length. Preferably, the contiguous stretch is included in the coding or non-coding strand of the reference nucleic acid sequence. Preferably, the contiguous stretch is in the coding or non-coding strand of the nucleic acid sequence of SEQ ID NO:26. The invention further provides an expression vector comprising such an antisense molecule.

The invention also provides a method of reducing GlyT-2 expression in a tissue or cell comprising applying to the tissue or cell a GlyT-2 expression reducing

- 8 -

effective amount of such an antisense molecule or a GlyT-2 expression reducing effective amount of an expression vector for expressing such an antisense molecule in a tissue or cell. Alternatively, the invention provides a method of treating a nervous system disorder or condition comprising applying to a tissue or cell of a human patient a nervous system disorder or condition treating effective amount of such an antisense molecule or a nervous system disorder or condition treating effective amount of an expression vector for expressing such an antisense molecule in a tissue or cell.

Further, the invention provides a method for detecting whether an animal has autoimmune antibodies against a glycine transporter, the method comprising contacting an antibody preparation from the animal or a body fluid from the animal with a polypeptide antigen comprising a glycine transporter or derived from the glycine transporter. Preferably, the polypeptide antigen comprises a contiguous sequence encoded by the protein sequence of SEQ ID NO:27 or with a sequence corresponding to the protein sequence of SEQ ID NO:27 except that it has one or more of the following substitutions (1) Gly¹⁰² to Ser, (2) Ser¹²⁴ to Phe, (3) Ile²⁷⁹ to Asn, (4) Arg³⁹³ to Gly, (5) Lys⁴⁵⁷ to Asn, (6) Asp⁴⁶³ to Asn, (7) Cys⁶¹⁰ to Tyr, (8) Ile⁶¹¹ to Val, (9) Phe⁷³³ to Ser, (10) Ile⁷³⁵ to Val, (11) Phe²⁴⁵ to Leu, (12) Val³⁰⁵ to Leu, (13) Thr³⁶⁶ to Ile or (14) Leu⁴⁰⁰ to Pro. Preferably, the contiguous sequence is at least about six amino acids in length, more preferably at least about ten amino acids in length, still more preferably at least about fifteen amino acids in length. In one embodiment of the invention, the peptide antigen is selective for antibodies against either a GlyT-1 transporter or a GlyT-2 transporter.

Brief Description of the Drawings

Figure 1 shows the alignment of several gene fragments of the human GlyT-2 gene.

Figure 2 illustrates which fragment clones were used to construct the clone incorporating the nucleic acid sequence of SEQ ID NO:20, a full-length clone of the human GlyT-2 gene.

Figure 3 shows a comparison between the nucleic acid sequence of SEQ ID NO:18 and the rat GlyT-2 sequence.

Figure 4 shows a comparison between the amino acid sequence of SEQ ID NO:19 and the rat GlyT-2 sequence.

Figure 5 shows the measurement of glycine transport in QT-6 cells either transfected with a human GlyT-2 expression vector or mock transfected.

Figure 6 shows the concentration dependence of glycine transport in QT-6

- 9 -

cells transfected with human GlyT-2.

Definitions

For the purposes of this application, the following terms shall have the meaning set forth below.

5 **○ Bioactive agent**

A bioactive agent is a substance such as a chemical that can act on a cell, virus, tissue, organ or organism, including but not limited to drugs (i.e. pharmaceuticals) to create a change in the functioning of the cell, virus, organ or organism. Preferably, the organism is a mammal, more preferably a human. In a preferred embodiment of the invention, the method of identifying bioactive agents of the invention is applied to organic molecules
10 having molecular weight of about 1500 or less.

○ extrinsically-derived nucleic acid

Extrinsically-derived nucleic acids are nucleic acids found in a cell that were introduced into the cell, a parent or ancestor of the cell, or a transgenic animal from which the cell
15 is derived through a recombinant technology.

○ extrinsic promoter functionally associated with a nucleic acid

An extrinsic promoter for a protein-encoding nucleic acid is a promoter distinct from that used in nature to express a nucleic acid for that protein. A promoter is functionally associated with the nucleic acid if in a cell that is compatible with the promoter the
20 promoter can act to allow the transcription of the nucleic acid.

○ nucleic acid-specific property

Nucleic acid-specific properties are properties that can be used to distinguish differing nucleic acid molecules. Such properties include, without limitation (i) the nucleotide
sequence of all or a portion of the molecule, (ii) the size of the molecule, for instance
25 determined by electrophoresis, (iii) the fragmentation pattern generated by treatment with chemicals that fragment nucleic acid or generated by nucleases and (iv) the ability of the molecule or fragments thereof to hybridize with defined nucleic acid probes or to generate amplicons with defined primers.

○ prospective agent

30 Prospective agents are substances which are being tested by the screening method of the invention to determine if they affect glycine transport.

○ Sequence identity

"Identity," as known in the art, is a relationship between two or more polypeptide sequences or two or more polynucleotide sequences, as determined by comparing the

- 10 -

sequences, particularly, as determined by the match between strings of such sequences. "Identity" is readily calculated by known methods (Computational Molecular Biology, Lesk, A.M., ed., Oxford University Press, New York, 1988; Biocomputing: Informatics and Genome Projects, Smith, D.W., ed., Academic Press, New York, 1993; Computer
5 Analysis of Sequence Data, Part I, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; Sequence Analysis in Molecular Biology, von Heinje, G., Academic Press, 1987; and Sequence Analysis Primer, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). While there exist a number of methods to measure identity between two sequences, the term is well known to skilled artisans (see, for
10 example, Sequence Analysis in Molecular Biology; Sequence Analysis Primer; and Carillo, H., and Lipman, D., SIAM J. Applied Math., 48: 1073 (1988)). Methods commonly employed to determine identity between sequences include, but are not limited to those disclosed in Carillo, H., and Lipman, D., SIAM J. Applied Math., 48:1073 (1988) or, preferably, in Needleman and Wunsch, J. Mol. Biol., 48: 443-445, 1970,
15 wherein the parameters are as set in version 2 of DNASIS (Hitachi Software Engineering Co., San Bruno, CA). Computer programs for determining identity are publicly available. Preferred computer program methods to determine identity between two sequences include, but are not limited to, GCG program package (Devereux, J., et al., Nucleic Acids Research 12(1): 387 (1984)), BLASTP, BLASTN, and FASTA (Atschul, S.F. et al., J.
20 Molec. Biol. 215: 403-410 (1990)). The BLAST X program is publicly available from NCBI (blast@ncbi.nlm.nih.gov) and other sources (BLAST Manual, Altschul, S., et al., NCBI NLM NIH Bethesda, MD 20894; Altschul, S., et al., J. Mol. Biol. 215: 403-410 (1990)).

Detailed Description of the Invention

25 The GlyT-2 nucleic acid sequence of SEQ ID NOS:18 and 26 or the corresponding encoded protein sequences of SEQ ID NOS:19 and 27, are human relatives of the rat GlyT-2 sequence reported in Liu et al., *J. Biol. Chem.* 268: 22802-22808, 1992. SEQ ID NO:21, the GlyT-2 protein sequence encoded by the nucleic acid sequence of SEQ ID NO:20, differs from the amino acid sequences of SEQ ID NOS:19 and 27, most
30 likely reflecting variant forms of human GlyT-2. Additional sequences set forth in SEQ IDs 1-34 reflect still further variations. These variations primarily arise from the use of cDNA from pooled mRNA for several donors to generate the clones. In total, the various human GlyT-2-derived nucleic acids that have been isolated reveal the following sequence variations:

- 11 -

	Nucleotide variations	Encoded Amino Acid Variations	Corresponding Amino Acid in Rat
	GAT ⁶ (from SEQ ID NOS:18 and 26) to GAC (from SEQ ID NO:3)	NONE (Asp ² to Asp)	Asp
5	A ³⁰⁴ GC (from SEQ ID NO:18) to GGC (from SEQ ID NOS:20 and 26)	Ser ¹⁰² to Gly	Ser
	CCC ³⁴² (from SEQ ID NOS:18 and 26) to CCG (from SEQ ID NO: 33)	NONE (Pro ¹¹⁴ to Pro)	Pro
10	C ³⁵² TG (from SEQ ID NOS:18 and 26) to TTG (from SEQ ID NO: 31)	NONE (Leu ¹¹⁸ to Leu)	Leu
15	TT ³⁷¹ T (from SEQ ID NO:20) to TCT (from SEQ ID NOS:18 and 26)	Phe ¹²⁴ to Ser	Ala
	C ⁵⁷¹ GA (from SEQ ID NOS:18 and 26) to TGA (from SEQ ID NO:7)	Arg ¹⁹¹ to STOP	Arg
20	T ⁷³³ TC (from SEQ ID NOS:18 and 26) to CTC (from SEQ ID NO: 31)	Phe ²⁴⁵ to Leu	Phe
	CCA ⁷⁷⁷ (from SEQ ID NOS:18 and 26) to CCG (from SEQ ID NO: 33)	NONE (Pro ²⁵⁹ to Pro)	Pro
25	AT ⁸³⁶ C (from SEQ ID NOS:18 and 26) to AAC (from SEQ ID NO:20)	Ile ²⁷⁹ to Asn	Ile
	G ⁹¹³ TA (from SEQ ID NOS:18 and 26) to CTA (from SEQ ID NO: 35)	Val ³⁰⁵ to Leu	Val
30	ACG ⁹⁵¹ (from SEQ ID NOS:18 and 26) to ACA (from SEQ ID NO: 29 and 31)	NONE (Thr ³¹⁷ to Thr)	Thr
35	AC ¹⁰⁹⁷ A (from SEQ ID NOS:18 and 26) to ATA (from SEQ ID NO: 31)	Thr ³⁶⁶ to Ile	Thr
	GAG ¹¹¹⁶ (from SEQ ID NO:20) to GAA (from SEQ ID NOS:18 and 26)	NONE (Glu ³⁷² to Glu)	Glu

- 12 -

	Nucleotide variations	Encoded Amino Acid Variations	Corresponding Amino Acid in Rat
	<u>G</u> ¹¹⁷⁷ GG (from SEQ ID NO:5) to <u>A</u> GG (from SEQ ID NOS:18 and 26)	Gly ³⁹³ to Arg	Arg
5	<u>CT</u> ¹¹⁹⁹ C (from SEQ ID NOS:18 and 26) to <u>CCC</u> (from SEQ ID NO: 33)	Leu ⁴⁰⁰ to Pro	Leu
	<u>AAC</u> ¹³⁷¹ (from SEQ ID NO:10) to <u>AAG</u> (from SEQ ID NOS:18 and 26)	Asn ⁴⁵⁷ to Lys	Lys
10	<u>G</u> ¹³⁸⁷ AT (from SEQ ID NOS:18 and 26) to <u>A</u> AT (from SEQ ID NO:12)	Asp ⁴⁶³ to Asn	Asp
	<u>TG</u> ¹⁸²⁹ C (from SEQ ID NOS:18 and 26) to <u>TAC</u> (from SEQ ID NO:22)	Cys ⁶¹⁰ to Tyr	Cys
15	<u>A</u> ¹⁸³¹ TT (from SEQ ID NOS:18 and 26) to <u>G</u> TT (from SEQ ID NO:20)	Ile ⁶¹¹ to Val	Ile
	<u>GAC</u> ²¹⁰³ (from SEQ ID NOS:18 and 26) to <u>GAA</u> (from SEQ ID NO:24)	NONE (Glu ⁷⁰¹ to Glu)	Glu
20	<u>TT</u> ²¹⁹⁸ T (from SEQ ID NOS:18 and 26) to <u>TCT</u> (from SEQ ID NO:24)	Phe ⁷³³ to Ser	Phe
25	<u>A</u> ²²⁰³ TA (from SEQ ID NOS:18 and 26) to <u>G</u> TA (from SEQ ID NO:22)	Ile ⁷³⁵ to Val	Ile

Irrespective of the source of this variation, the point variations in peptide sequence, excepting the insertion of the stop codon, are believed not to adversely affect the functioning of GlyT-2. The GlyT-2 protein sequence of SEQ ID NO:19 and SEQ ID NO: 27 are especially most preferred, with SEQ ID NO: 27 most preferred. The nucleic acid sequence of SEQ ID NO:26 is believed to represent the major consensus sequence.

The above-described variations primarily reflect sequence variations between human individuals. The material used to generate the nucleic acid sequences described above comprised pools from either twenty-six or ninety-two individuals, depending on the particular nucleic acid sequence. The use of pooled source material, together with the

- 13 -

prevalence of silent or conservative substitutions, support the conclusion that the variations are reflective of human-derived variations rather than mutations generated by the amplification reactions.

5 The relationship between the human nucleotide sequence of SEQ ID NO:18 and the rat nucleotide sequence for GlyT-2, and between the protein sequences that they encode, is as set forth in the tables below. The relatedness values set forth in these tables was determined using the FASTA computer program described by Pearson and Lipman, *Proc. Natl. Acad. Sci. USA* 85: 2444-2448, 1988.

10

Nucleotide Sequence (numbered as in SEQ ID NO:18)	Percent Identity
nt 1-2397 (whole sequence)	89
nt 1-600	82.5
nt 60-170	78
nt 600-2397	91.2

15

Amino Acid Sequence (numbered as in SEQ ID NO:19)	Percent Identity
aa 1-797	94.4
aa 1-150	77.1
aa 1-200	80.3
aa 150-797	98.5
aa 200-797	99.2

20

Nucleic Acid - encoding glycine transporter

25

To construct non-naturally occurring glycine transporter-encoding nucleic acids, the native sequences can be used as a starting point and modified to suit particular needs. For instance, the sequences can be mutated to incorporate useful restriction sites. See Maniatis et al. *Molecular Cloning, a Laboratory Manual* (Cold Spring Harbor Press,

- 14 -

1989). Such restriction sites can be used to create "cassettes", or regions of nucleic acid sequence that are readily substituted using restriction enzymes and ligation reactions. The cassettes can be used to substitute synthetic sequences encoding mutated glycine transporter amino acid sequences. Alternatively, the glycine transporter-encoding sequence can be substantially or fully synthetic. See, for example, Goeddel et al., *Proc. Natl. Acad. Sci. USA*, 76, 106-110, 1979. For recombinant expression purposes, codon usage preferences for the organism in which such a nucleic acid is to be expressed are advantageously considered in designing a synthetic glycine transporter-encoding nucleic acid. For example, a nucleic acid sequence incorporating prokaryotic codon preferences can be designed from a mammalian-derived sequence using a software program such as Oligo-4, available from National Biosciences, Inc. (Plymouth, MN).

The nucleic acid sequence embodiments of the invention are preferably deoxyribonucleic acid sequences, preferably double-stranded deoxyribonucleic acid sequences. However, they can also be ribonucleic acid sequences.

Numerous methods are known to delete sequence from or mutate nucleic acid sequences that encode a protein and to confirm the function of the proteins encoded by these deleted or mutated sequences. Accordingly, the invention also relates to a mutated or deleted version of a human nucleic acid sequence that encodes a protein that retains the ability to bind specifically to glycine and to transport glycine across a membrane. These analogs can have N-terminal, C-terminal or internal deletions, so long as GlyT-2 function is retained. The remaining human GlyT-2 protein sequence will preferably have no more than about 4 amino acid variations, preferably no more than 2 amino acid variations, more preferably no more than 1 amino acid variation, relative to the protein sequence of SEQ ID NO:27 or with a sequence corresponding to the protein sequence of SEQ ID NO:27 except that it has one or more of the following substitutions (1) Gly¹⁰² to Ser, (2) Ser¹²⁴ to Phe, (3) Ile²⁷⁹ to Asn, (4) Arg³⁹³ to Gly, (5) Lys⁴⁵⁷ to Asn, (6) Asp⁴⁶³ to Asn, (7) Cys⁶¹⁰ to Tyr, (8) Ile⁶¹¹ to Val, (9) Phe⁷³³ to Ser, (10) Ile⁷³⁵ to Val, (11) Phe²⁴⁵ to Leu, (12) Val³⁰⁵ to Leu, (13) Thr³⁶⁶ to Ile, or (14) Leu⁴⁰⁰ to Pro. More preferably, the variations are relative to the protein sequence of SEQ ID NOS:19 or 27, still more preferably SEQ ID NO:27. In one preferred embodiment, the protein embodiments of the invention are defined relative to the protein sequence of SEQ ID NO:27 or with a sequence corresponding to the protein sequence of SEQ ID NO:27 except that it has one or more of the following substitutions (1) Gly¹⁰² to Ser, (2) Ser¹²⁴ to Phe, (3) Ile²⁷⁹ to Asn, (4) Arg³⁹³ to Gly, (5) Lys⁴⁵⁷ to Asn, (6) Asp⁴⁶³ to Asn, (7)

- 15 -

Cys⁶¹⁰ to Tyr, (8) Ile⁶¹¹ to Val, (9) Phe⁷³³ to Ser, or (10) Ile⁷³⁵ to Val. The point variations are preferably conservative point variations. Preferably, the analogs will have at least about 96% sequence identity, preferably at least about 97%, more preferably at least about 98%, still more preferably at least about 99%, yet still more preferably at least about 99.5%, to the protein sequence of SEQ ID NO:27 or with a sequence corresponding to the protein sequence of SEQ ID NO:27 except that it has one or more of the following substitutions (1) Gly¹⁰² to Ser, (2) Ser¹²⁴ to Phe, (3) Ile²⁷⁹ to Asn, (4) Arg³⁹³ to Gly, (5) Lys⁴⁵⁷ to Asn, (6) Asp⁴⁶³ to Asn, (7) Cys⁶¹⁰ to Tyr, (8) Ile⁶¹¹ to Val, (9) Phe⁷³³ to Ser, (10) Ile⁷³⁵ to Val, (11) Phe²⁴⁵ to Leu, (12) Val³⁰⁵ to Leu, (13) Thr³⁶⁶ to Ile or (14) Leu⁴⁰⁰ to Pro. More preferably, the variations are relative to the protein sequence of SEQ ID NOS:19 or 27, still more preferably SEQ ID NO:27.

Mutational and deletional approaches can be applied to all of the nucleic acid sequences of the invention that express human GlyT-2 proteins. As discussed above, conservative mutations are preferred. Such conservative mutations include mutations that switch one amino acid for another within one of the following groups:

1. Small aliphatic, nonpolar or slightly polar residues: Ala, Ser, Thr, Pro and Gly;
2. Polar, negatively charged residues and their amides: Asp, Asn, Glu and Gln;
3. Polar, positively charged residues: His, Arg and Lys;
4. Large aliphatic, nonpolar residues: Met, Leu, Ile, Val and Cys; and
5. Aromatic residues: Phe, Tyr and Trp.

A preferred listing of conservative variations is the following:

Original Residue	Variation
Ala	Gly, Ser
Arg	Lys
Asn	Gln, His
Asp	Glu
Cys	Ser
Gln	Asn
Glu	Asp

- 16 -

Original Residue	Variation
Gly	Ala, Pro
His	Asn, Gln
Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Tyr, Ile
Phe	Met, Leu, Tyr
Ser	Thr
Thr	Ser
Trp	Tyr
Tyr	Trp, Phe
Val	Ile, Leu

The types of variations selected may be based on the analysis of the frequencies of amino acid variations between homologous proteins of different species developed by Schulz et al., *Principles of Protein Structure*, Springer-Verlag, 1978, on the analyses of structure-forming potentials developed by Chou and Fasman, *Biochemistry* 13, 211, 1974 and *Adv. Enzymol.* 47, 45-149, 1978, and on the analysis of hydrophobicity patterns in proteins developed by Eisenberg et al., *Proc. Natl. Acad. Sci. USA* 81, 140-144, 1984; Kyte & Doolittle; *J. Molec. Biol.* 157, 105-132, 1981, and Goldman et al., *Ann. Rev. Biophys. Chem.* 15, 321-353, 1986. All of the references of this paragraph are incorporated herein in their entirety by reference.

Since the ten identified point variations which create amino acid substitutions between the various human GlyT-2 mRNAs identified herein are believed to be useful in creating functional GlyT-2, proteins incorporating all combinations of these point variations are believed to be functional. These variations are within the invention.

For the purposes of this application, a nucleic acid of the invention is "isolated" if it has been separated from other macromolecules of the cell or tissue from which it is derived. Preferably, the composition containing the nucleic acid is at least about 10-fold enriched, with respect to nucleic acid content, over the composition of the source cells. Preferably, the nucleic acid is substantially pure, meaning purity of at least about 60% w/w with respect to other nucleic acids, more preferably about 80%, still

- 17 -

more preferably about 90%, yet more preferably about 95%.

Hybridization Probes

It will be recognized that many deletional or mutational analogs of nucleic acid sequences for a glycine transporter will be effective hybridization probes for glycine transporter-encoding nucleic acid. Accordingly, the invention relates to nucleic acid sequences that hybridize with such glycine transporter-encoding nucleic acid sequences under stringent conditions. Preferably, the nucleic acid sequence hybridizes with the nucleic acid sequence of SEQ ID NO:26 or with a nucleic acid sequence that varies therefrom by one or more of the following substitutions (a) T⁶ to C, (b) G³⁰⁴ to A, (c) C³⁷¹ to T, (d) C⁵⁷¹ to T, (e) T⁸³⁶ to A, (f) A¹¹¹⁶ to G, (g) A¹¹⁷⁷ to G, (h) G¹³⁷¹ to C, (i) G¹³⁸⁷ to A, (j) G¹⁸²⁹ to A, (k) A¹⁸³¹ to G, (l) G²¹⁰³ to A, (m) T²¹⁹⁸ to C, (n) A²²⁰³ to G, (o) C³⁴² to G, (p) C³⁵² to T, (q) T⁷³³ to C, (r) A⁷⁷⁷ to G, (s) G⁹¹³ to C, (t) G⁹⁵¹ to A, (u) C¹⁰⁹⁷ to T or (v) T¹¹⁹⁹ to C. In one embodiment, the nucleic acid (or the functional equivalent) embodiments of the invention are defined relative to the nucleic acid sequence of SEQ ID NO:26 or with a nucleic acid sequence that varies therefrom by one or more of the following substitutions (a) T⁶ to C, (b) G³⁰⁴ to A, (c) C³⁷¹ to T, (d) C⁵⁷¹ to T, (e) T⁸³⁶ to A, (f) A¹¹¹⁶ to G, (g) A¹¹⁷⁷ to G, (h) G¹³⁷¹ to C, (i) G¹³⁸⁷ to A, (j) G¹⁸²⁹ to A, (k) A¹⁸³¹ to G, (l) G²¹⁰³ to A, (m) T²¹⁹⁸ to C, or (n) A²²⁰³ to G.

"Stringent conditions" refers to conditions that allow for the hybridization of substantially related nucleic acid sequences. For instance, such conditions will generally allow hybridization of sequence with at least about 85% sequence identity, preferably with at least about 90% sequence identity, more preferably with at least about 95% sequence identity. Such hybridization conditions are described by Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Press, 1989.

Hybridization conditions and probes can be adjusted in well-characterized ways to achieve selective hybridization of human-derived probes.

Nucleic acid molecules that will hybridize to a glycine transporter-encoding nucleic acid under stringent conditions can be identified functionally, using methods outlined above, or by using for example the hybridization rules reviewed in Sambrook et al., *Molecular Cloning: A Laboratory Manual*, 2nd ed., Cold Spring Harbor Press, 1989.

Without limitation, examples of the uses for hybridization probes include : histochemical uses such as identifying tissues that express the human GlyT-2 transporter; measuring mRNA levels, for instance to identify a sample's tissue type or to identify cells that express abnormal levels of glycine transporter; and detecting polymorphisms in the

- 18 -

glycine transporter gene. RNA hybridization procedures are described in Maniatis et al. *Molecular Cloning, a Laboratory Manual* (Cold Spring Harbor Press, 1989).

PCR Primers

Rules for designing polymerase chain reaction ("PCR") primers are now established, as reviewed by *PCR Protocols*, Cold Spring Harbor Press, 1991. Degenerate primers, i.e., preparations of primers that are heterogeneous at given sequence locations, can be designed to amplify nucleic acid sequences that are highly homologous to, but not identical to, a human GlyT-2 nucleic acid. Strategies are now available that allow for only one of the primers to be required to specifically hybridize with a known sequence. See, Froman et al., *Proc. Natl. Acad. Sci. USA* 85: 8998, 1988 and Loh et al. *Science* 243: 217, 1989. For example, appropriate nucleic acid primers can be ligated to the nucleic acid sought to be amplified to provide the hybridization partner for one of the primers. In this way, only one of the primers need be based on the sequence of the nucleic acid sought to be amplified.

PCR methods of amplifying nucleic acid will utilize at least two primers. One of these primers will be capable of hybridizing to a first strand of the nucleic acid to be amplified and of priming enzyme-driven nucleic acid synthesis in a first direction. The other will be capable of hybridizing the reciprocal sequence of the first strand (if the sequence to be amplified is single stranded, this sequence will initially be hypothetical, but will be synthesized in the first amplification cycle) and of priming nucleic acid synthesis from that strand in the direction opposite the first direction and towards the site of hybridization for the first primer. Conditions for conducting such amplifications, particularly under preferred stringent hybridization conditions, are well known. See, for example, *PCR Protocols*, Cold Spring Harbor Press, 1991.

Vectors

A suitable expression vector is capable of fostering expression of the included GlyT-2 encoding DNA in a host cell, which can be eukaryotic, fungal, or prokaryotic. Suitable expression vectors include pRc/CMV (Invitrogen, San Diego, CA), pRc/RSV (Invitrogen), pcDNA3 (Invitrogen), Zap Express Vector (Stratagene Cloning Systems, LaJolla, CA); pBk/CMV or pBk-RSV vectors (Stratagene), Bluescript II SK +/- Phagemid Vectors (Stratagene), LacSwitch (Stratagene), pMAM and pMAM neo (Clontech, Palo Alto, CA), pKSV10 (Pharmacia, Piscataway, NJ), pCRscript (Stratagene) and pCR2.1 (Invitrogen), among others. Useful yeast expression systems include, for example, pYEUra3 (Clontech). Useful baculovirus vectors include several viral vectors

- 19 -

from Invitrogen (San Diego, CA) such as pVL1393, pVL1392, pBluBac2, pBluBacHis A, B or C, and pbacPAC6 (from Clontech).

Cells

5 In one embodiment of the invention, the transporter is preferably expressed in a mammalian cell line, preferably a transformed cell line with an established cell culture history. In this embodiment, particularly preferred cell lines include COS-1, COS-7, LM(tk⁻), HeLa, HEK293, CHO, Rat-1 and NIH3T3. Other preferred cells include avian cells such as QT-6 cells. Other cells that can be used include insect cells such as drosophila cells, fish cells, amphibian cells and reptilian cells.

10 In another embodiment, the transporter is expressed in a cell line that is more inexpensively maintained and grown than are mammalian cell lines, such as a bacterial cell line or a yeast cell line.

Isolated Glycine Transporter

15 The invention also provides for the human GlyT-2 proteins encoded by any of the nucleic acids of the invention preferably in a purity of at least about 80% with respect to proteins, preferably 90%, more preferably 95%. The purities are achieved, for example, by applying protein purification methods, such as those described below, to a lysate of a recombinant cell according to the invention.

20 The human GlyT-2 variants of the above paragraphs can be used to create organisms or cells that produce human GlyT-2 activity. Purification methods, including associated molecular biology methods, are described below.

Method of Producing Glycine Transporter

One simplified method of isolating polypeptides synthesized by an organism under the direction of one of the nucleic acids of the invention is to recombinantly
25 express a fusion protein wherein the fusion partner is readily affinity purified. For instance, the fusion partner can be glutathione S-transferase, which is encoded on commercial expression vectors (e.g., vector pGEX4T3, available from Pharmacia, Piscataway, NJ). The fusion protein can then be purified on a glutathione affinity column (for instance, that available from Pharmacia, Piscataway, New Jersey).
30 Additional fusion partners are available for example in various expression vectors sold by Invitrogen (Carlsbad, CA). Of course, the recombinant polypeptides can be affinity purified without such a fusion partner using an appropriate antibody that binds to GlyT-2. Methods of producing such antibodies are available to those of ordinary skill in light of the ample description herein of GlyT-2 expression systems and known antibody

- 20 -

production methods. See, for example, Ausubel et al., *Short Protocols in Molecular Biology*, John Wiley & Sons, New York, 1992. If fusion proteins are used, the fusion partner can be removed by partial proteolytic digestion approaches that preferentially attack unstructured regions such as the linkers between the fusion partner and GlyT-2.

- 5 The linkers can be designed to lack structure, for instance using the rules for secondary structure forming potential developed, for instance, by Chou and Fasman, *Biochemistry* 13, 211, 1974 and Chou and Fasman, *Adv. in Enzymol.* 47, 45-147, 1978. The linker can also be designed to incorporate protease target amino acids, such as, arginine and lysine residues, the amino acids that define the sites cleaved by trypsin, or such as a target
- 10 sequence for enterokinase, for example AspAspAspAspLys, which is cleaved after the lysine residue. To create the linkers, standard synthetic approaches for making oligonucleotides can be employed together with standard subcloning methodologies. Other fusion partners besides GST can be used. Procedures that utilize eukaryotic cells, particularly mammalian cells, are preferred since these cells will post-translationally
- 15 modify the protein to create molecules highly similar to or functionally identical to native proteins.

- Additional purification techniques can be applied, including without limitation, preparative electrophoresis, FPLC (Pharmacia, Uppsala, Sweden), HPLC (e.g., using gel filtration, reverse-phase or mildly hydrophobic columns), gel filtration,
- 20 differential precipitation (for instance, "salting out" precipitations), ion-exchange chromatography and affinity chromatography.

- Because GlyT-2 is a membrane protein, which by analogy to related transporter proteins is believed to have twelve transmembrane sequences, isolation methods will often utilize detergents, generally non-ionic detergents, to maintain the
- 25 appropriate secondary and tertiary structure of the protein. See, for example, Lopez-Corcuera et al., *J. Biol. Chem.* 266: 24809-24814, 1991. For a description of methods for re-integrating a solubilized transporter into a membrane, see Lopez-Corcuera et al., *J. Biol. Chem.* 266: 24809-24814, 1991.

- The isolation of GlyT-2 can comprise isolating membranes from cells that
- 30 have been transformed to express GlyT-2. Preferably, such cells express GlyT-2 in sufficient copy number such that the amount of GlyT-2 in a membrane fraction is at least about 10-fold higher than that found in comparable membranes from cells that naturally express GlyT-2, more preferably the amount is at least about 100-fold higher.

Preferably, the protein is substantially pure, meaning a purity of at least 60%

- 21 -

wt/wt with respect to other proteins. For the purposes of this application, GlyT-2 is "isolated" if it has been separated from other proteins or other macromolecules of the cell or tissue from which it is derived. Preferably, the composition containing GlyT-2 is at least about 10-fold enriched, preferably at least about 100-fold, with respect to protein content, over the composition of the source cells.

Expression of GlyT-2 by RNA Insertion

It will be recognized that human GlyT-2 can be expressed by the simple method of inserting mRNA into a cell. RNA for these uses can be prepared by sub-cloning the nucleic acid encoding a protein with GlyT-2 activity into a vector containing a promoter for high efficiency *in vitro* transcription, such as a SP6 or T7 RNA polymerase promoter. RNA production from the vector can be conducted, for instance, with the method described in Ausubel et al., *Short Protocols in Molecular Biology*, John Wiley & Sons, New York, 1992, pp. 10-63 to 10-65. Insertion of RNA into *Xenopus*-derived oocytes is described, for instance, in Liu et al. *FEBS Letters* 305: 110-114, 1992 and Bannon et al., *J. Neurochem.* 54: 706-708, 1990.

Alternatively, it will be recognized that human GlyT-2 can be expressed by the simple method of inserting mRNA into an *in vitro* translation system, which can be a membrane-containing translation system. Expression of proteins *in vitro* is described, for instance, in Ausubel et al., *Short Protocols in Molecular Biology*, John Wiley & Sons, New York, 1992, pp. 10-63 to 10-65. See, also, Guastella et al., *Science* 249: 1303-1306, 1990 (*in vitro* expression of a transporter). The use of subcellular membranous material to produce membrane proteins *in vitro* is described in Walter and Blobel, *Meth. Enzymol.* 96: 84, 1983 (for rabbit reticulocyte translation system) and Spiess and Lodish, *Cell* 44: 177, 1986 (for wheat-germ translation system).

Method of Characterizing or Identifying agent

A method for the analysis of or screening for a bioactive agent for treatment of a disease or condition associated with a nervous system disorder or condition comprises culturing separately first and second cells, wherein the first and second cells are preferably of the same species, more preferably of the same strain thereof, and comprise an exogenous nucleic acid encoding a glycine transporter as described herein. The nervous system disorders or conditions for which the agent can be used for treatment include, but are not limited to, (a) pain, (b) myoclonus, (c) muscle spasm, (d) muscle hyperactivity, (e) epilepsy or (f) spasticity such as that associated with stroke, head trauma, neuronal cell death, multiple sclerosis, spinal cord injury, dystonia, Huntington's

- 22 -

disease or amyotrophic lateral sclerosis. In this method, the first cell is contacted with the bioactive agent or a prospective agent, which is preferably a compound, such as a peptide or an organic compound in the presence of glycine, which preferably incorporates a radioisotope, such as ^3H or ^{14}C . The contacted first cell is then tested for enhancement
5 or inhibition of glycine transport into the first cell as compared to glycine transport into the second cell that was not contacted with the compound (i.e., the control cell). Such analysis or screening preferably includes activities of finding, learning, discovering, determining, identifying, or ascertaining.

Alternatively, the assay can utilize a composition comprising an isolated
10 GlyT-2 transporter in place of cells. Preferably, such preparation of isolated transporter will comprise membrane or lipid bilayer, preferably in vesicles, which vesicles have an inside and an outside across which transport can be measured. See, for example, Kanner, *Biochemistry* 17: 1207-1211, 1978.

A bioactive agent is an enhancer of glycine transport uptake if at the end of
15 the test the amount of intracellular, intravesicle or otherwise transported glycine is greater in the agent-contacted composition than in the non-agent-contacted composition; conversely, a bioactive agent is an inhibitor of glycine transport if the amount of intracellular or intravesicle glycine is greater in the non-agent-contacted composition as compared to the other. Preferably, the difference in glycine uptake between a tested first
20 composition and a control second composition is at least about two-fold; more preferably, the difference is at least about five-fold; most preferably, the difference is at least about ten-fold or greater.

A bioactive agent that is an inhibitor or an enhancer with respect to the
GlyT-2 transporter may have a neutral or opposite effect with another glycine transporter,
25 such as one of the GlyT-1 transporters. Preferred bioactive agents have specificity to enhance or inhibit the GlyT-2 transporter and have neutral or negligible effect on other glycine transporters. Preferably, a bioactive agent has at least an order of magnitude greater potency, reflected in a concentration dependent parameter such as the IC_{50} value, in inhibiting or activating glycine uptake mediated by the GlyT-2 transporter as compared
30 to its effect on the second glycine transporter. More preferred agents have greater potencies of at least about 100-fold for one of the glycine transporters as compared to the other.

The bioactive agent can be any compound, material, composition, mixture, or chemical, that can be presented to a glycine transporter in a form that allows for the

- 23 -

agent to diffuse so as to contact the transporter. Such bioactive agents include but are not limited to polypeptides preferably of two up to about 25 amino acids in length, more preferably from two to about ten, yet more preferably from two to about five amino acids in length. Other suitable bioactive agents in the context of the present invention include
5 small organic compounds, preferably of molecular weight between about 100 daltons and about 5,000 daltons, and are composed of such functionalities as alkyl, aryl, alkene, alkyne, halo, cyano and other groups, including heteroatoms or not. Such organic compounds can be carbohydrates, including simple sugars, amino or imino acids, nucleic acids, steroids, and others. The chemicals tested as prospective agents can be prepared
10 using combinatorial chemical processes known in the art or conventional means for chemical synthesis. Preferably, bioactive agents are useful as drugs for treatment of nervous system disorders or conditions.

Some compounds that inhibit GlyT-1 or GlyT-2 mediated transport also bind to the glycine binding site on the strychnine-sensitive receptor, or to the glycine binding
15 site on the NMDA receptor. Such binding to the strychnine-sensitive receptor can be identified by a binding assay whereby, for example, radiolabeled strychnine is placed in contact with a preparation of strychnine-sensitive receptors, such as can be prepared from a membrane fraction from spinal cord or brain stem tissue. A membrane fraction can be prepared using conventional means, including, for example, methods of homogenization
20 and centrifugation.

Such binding to the NMDA receptor can be identified by a binding assay whereby, for example, radiolabeled glycine is placed in contact with a preparation of
NMDA receptors, such as can be prepared from a membrane fraction from neuronal cells or brain tissue. Grimwood et al., *Molec. Pharmacol.*, 41:923-930, 1992. The NMDA
25 receptors located in such membranes are treated using mild detergent, such as about 0.1% to about 0.5% saponin, to remove any endogenous glycine or glutamate.

The ligand used in such a binding assay is radiolabeled with any detectable isotope, such as radioactive isotopes of carbon or hydrogen. Specific binding of the radiolabeled ligand is then determined by subtracting the radioactivity due to non-specific
30 binding from that which is due to total (*i.e.*, specific and non-specific) binding of the radiolabeled ligand. The radioactivity due to non-specific binding is determined by measuring the amount of radiolabel associated with a strychnine-sensitive or NMDA receptor-containing membrane fraction that has been contacted with both radiolabeled ligand and a significant excess of non-radiolabeled ligand, such as a 100-fold excess.

- 24 -

The radioactivity due to total binding of the radiolabeled ligand is determined by measuring the amount of radiolabel bound to the receptor preparation in the absence of non-radiolabeled ligand. For the NMDA receptor, one can also measure binding to the glycine site on the receptor using labeled analogs of amino acids, such as, for example, dichlorokynurenic acid or L-689,560. See, for example, Grimwood et al., *Molecular Pharmacol.*, 49: 923-930, 1992.

Functional ion-flux assays are used to measure the effect of compounds identified by the present invention in enhancing or inhibiting calcium flux (for NMDA receptor preparations) or chloride flux (for strychnine-sensitive receptor preparations).

10 This test is performed on cell cultures that have membrane-bound NMDA receptors or strychnine-sensitive receptors and glycine transporters. Such cells include neuronal cells generally, including those of the brain stem and spinal cord, and cell lines derived therefrom, and any other cell that has been induced or transfected to express NMDA receptors or strychnine-sensitive receptors. Calcium used in such a test is commonly the

15 ⁴⁵Ca isotope, although other calcium measuring techniques can be used as well, such as calcium-associated fluorescence, which can be fluorescence associated with a calcium chelator, and the like. Chloride used in such a test usually includes the isotope ³⁶Cl. By whatever method the calcium or chloride is monitored, ion flux can be enhanced or inhibited as a result of the discrete addition of a bioactive agent of the present invention.

20 An advantage of this system is that it allows one to monitor the net effect on NMDA receptor or strychnine-sensitive receptor function of a compound that interacts with both the glycine site on a receptor and on a glycine transporter.

GlyT-2 inhibitors that are also strychnine-sensitive receptor agonists act in the above-described indications by increasing glycine concentrations at the strychnine-sensitive receptor-expressing synapses via inhibition of the glycine transporter, and via directly enhancing strychnine-sensitive receptor activity. Glycine transporter inhibitors that are also strychnine-sensitive receptor antagonists can nonetheless retain activity in treating these indications, for example if the increase in glycine due to glycine transport inhibition prevails over the strychnine-sensitive receptor antagonism. Where the

25 strychnine-sensitive receptor antagonist activity prevails over the effect of increased extracellular glycine resulting from inhibition of the glycine transporter, these compounds are useful in treating conditions associated with decreased muscle activity such as myasthenia gravis.

30

As discussed above, the bioactive agents of the invention can have a number

- 25 -

of pharmacological actions. The relative effectiveness of the compounds can be assessed in a number of ways, including the following:

1. Comparing the activity mediated through GlyT-1 and GlyT-2 transporters. This testing identifies bioactive agents (a) that are more active against GlyT-1 transporters and thus more useful in treating or preventing schizophrenia, increasing cognition and enhancing memory or (b) that are more active against GlyT-2 transporters and thus more useful in treating or preventing epilepsy, pain or spasticity.
2. Testing for strychnine-sensitive receptor or NMDA receptor binding. This test establishes whether there is sufficient binding at this site to warrant further examination of the pharmacological effect of such binding.
3. Testing the activity of the compounds in enhancing or diminishing ion fluxes in primary tissue culture, for example chloride ion fluxes mediated by strychnine-sensitive receptors or calcium ion fluxes mediated by NMDA receptors. A bioactive agent that increases ion flux either (a) has little or no antagonist activity at the strychnine-sensitive receptor and should not affect the potentiation of glycine activity through GlyT-2 transporter inhibition or (b), if marked increases are observed over results with comparative GlyT-2 inhibitors that have little direct interaction with strychnine-sensitive receptors, then the agent is a receptor agonist.

In some cases, the agent analysis method of the invention will be used to characterize whether a bioactive agent is useful in treating an indication in which NMDA receptors and GlyT-1 transporters are implicated. In this case, generally, a lower measure of activity with respect to strychnine-sensitive receptors and GlyT-2 transporters is more desirable.

Antisense Therapies

One aspect of the present invention is directed to the use of "antisense" nucleic acid to treat neurological indications such as those identified above. The approach involves the use of an antisense molecule designed to bind mRNA coding for a GlyT-2, thereby stopping or inhibiting the translation of the mRNA, or to bind to the GlyT-2 gene to interfere with its transcription. For discussion of the design of nucleotide sequences that bind genomic DNA to interfere with transcription, see Helene, *Anti-Cancer Drug Design* 6, 569, 1991. Once the sequence of the mRNA sought to be bound is known, an antisense molecule can be designed that binds the sense strand by the Watson-Crick base-pairing rules, forming a duplex structure analogous to the DNA double helix. *Gene Regulation: Biology of Antisense RNA and DNA*, Erikson and Ixant,

- 26 -

eds., Raven Press, New York, 1991; Helene, *Anti-Cancer Drug Design*, 6:569 (1991); Crooke, *Anti-Cancer Drug Design* 6, 609, 1991.

A serious barrier to fully exploiting antisense technology is the problem of efficiently introducing into cells a sufficient number of antisense molecules to effectively
5 interfere with the translation of the targeted mRNA or the function of DNA. One method that has been employed to overcome this problem is to covalently modify the 5' or the 3' end of the antisense polynucleic acid molecule with hydrophobic substituents. These modified nucleic acids generally gain access to the cells interior with greater efficiency. See, for example, Bourtin et al., *FEBS Lett.* 23,1382-1390, 1989; Shea et al, *Nucleic
10 Acids Res.* 18, 3777-3783, 1990. Additionally, the phosphate backbone of the antisense molecules has been modified to remove or diminish negative charge (see, for example, Agris et al., *Biochemistry* 25, 6268, 1986; Cazenave and Helene in *Antisense Nucleic Acids and Proteins: Fundamentals and Applications*, Mol and Van der Krol, eds., p. 47 et seq., Marcel Dekker, New York, 1991) or the purine or pyrimidine bases have been
15 modified (see, for example, *Antisense Nucleic Acids and Proteins: Fundamentals and Applications*, Mol and Van der Krol, eds., p. 47 et seq., Marcel Dekker, New York, 1991; Milligan et al. in *Gene Therapy For Neoplastic Diseases*, Huber and Laso, eds., p. 228 et seq., New York Academy of Sciences, New York, 1994). Other methods to overcome the cell penetration barrier include incorporating antisense polynucleic acid
20 sequences into expression vectors that can be inserted into the cell in low copy number, but which in the cell can direct the cellular machinery to synthesize more substantial amounts of antisense polynucleic molecules. See, for example, Farhood et al., *Ann. N.Y. Acad. Sci.* 716, 23, 1994. This strategy includes the use of recombinant viruses that have an expression site into which the antisense sequence has been incorporated. See, e.g.,
25 Boris-Lawrie and Temin, *Ann. N.Y. Acad. Sci.*, 716:59 (1994). Others have tried to increase membrane permeability by neutralizing the negative charges on antisense molecules or other nucleic acid molecules with polycations. See, e.g. Wu and Wu, *Biochemistry*, 27:887-892, 1988; Behr et al., *Proc. Natl. Acad Sci U.S.A.* 86:6982-6986, 1989.

30 For gene therapy such as antisense therapy, medical workers often try to incorporate, into one or more cell types of an organism, a DNA vector capable of directing the synthesis of a protein missing from the cell or useful to the cell or organism when expressed in greater amounts. The methods for introducing DNA to cause a cell to produce a new protein or a greater amount of a protein are called "transfection" methods.

- 27 -

See, generally, *Neoplastic Diseases*, Huber and Lazo, eds., New York Academy of Science, New York, 1994; Feigner, *Adv. Drug Deliv. Rev.*, 5:163 (1990); McLachlin, et al., *Progr. Nucl. Acids Res. Mol. Biol.*, 38:91 (1990); Karlsson, S. *Blond*, 78:2481 (1991); Einerhand and Valerio, *Curr. Top. Microbiol. Immunol.*, 177:217-235 (1992); Makdisi et al., *Prog. Liver Dis.*, 10:1 (1992); Litzinger and Huang, *Biochim. Biophys. Acta*, 1113:201 (1992); Morsy et al., *J.A.M.A.*, 270:2338 (1993); Dorudi et al., *British J. Surgery*, 80:566 (1993).

Other general methods of incorporating nucleic acids into cells include calcium phosphate precipitation of nucleic acid and incubation with the target cells (Graham and Van der Eb, *Virology*, 52:456, 1983), co-incubation of nucleic acid, DEAE-dextran and cells (Sompayrac and Danna, *Proc. Natl. Acad. Sci.*, 12:7575, 1981), electroporation of cells in the presence of nucleic acid (Potter et al., *Proc. Natl. Acad. Sci.*, 81:7161-7165, 1984), incorporating nucleic acid into virus coats to create transfection vehicles (Gitman et al., *Proc. Natl. Acad. Sci. U.S.A.*, 82:7309-7313, 1985) and incubating cells with nucleic acid incorporated into liposomes (Wang and Huang, *Proc. Natl. Acad. Sci.*, 84:7851-7855, 1987). One approach to gene therapy is to incorporate the gene sought to be introduced into the cell into a virus, such as a herpes virus, adenovirus, parvovirus or a retrovirus. See, for instance, Akli et al., *Nature Genetics* 3, 224, 1993.

The nucleic acid compositions of the invention can be, for example, administered orally, topically, rectally, nasally, vaginally, by inhalation, for example by use of an aerosol, or parenterally, e.g. intramuscularly, subcutaneously, intraperitoneally, intraventricularly, or intravenously. The nucleic acid compositions can be administered alone, or they can be combined with a pharmaceutically-acceptable carrier or excipient according to standard pharmaceutical practice. For the oral mode of administration, the nucleic acid compositions can be used in the form of tablets, capsules, lozenges, troches, powders, syrups, elixirs, aqueous solutions and suspensions, and the like. In the case of tablets, carriers that can be used include lactose, sodium citrate and salts of phosphoric acid. Various disintegrants such as starch, and lubricating agents such as magnesium stearate, sodium lauryl sulfate and talc, are commonly used in tablets. For oral administration in capsule form, useful diluents are lactose and high molecular weight polyethylene glycols. When aqueous suspensions are required for oral use, the nucleic acid compositions can be combined with emulsifying and suspending agents. If desired, certain sweetening and/or flavoring agents can be added. For parenteral administration,

- 28 -

sterile solutions of the conjugate are usually prepared, and the pH of the solutions are suitably adjusted and buffered. For intravenous use, the total concentration of solutes should be controlled to render the preparation isotonic. For ocular administration, ointments or droppable liquids may be delivered by ocular delivery systems known to the art such as applicators or eye droppers. Such compositions can include mucomimetics such as hyaluronic acid, chondroitin sulfate, hydroxypropyl methylcellulose or poly(vinyl alcohol), preservatives such as sorbic acid, EDTA or benzylchonium chloride, and the usual quantities of diluents and/or carriers. For pulmonary administration, diluents and/or carriers will be selected to be appropriate to allow the formation of an aerosol.

Generally, the nucleic acid compositions will be administered in an effective amount. For pharmaceutical uses, an effective amount is an amount effective to either (1) reduce the symptoms of the indication sought to be treated or (2) induce a pharmacological change relevant to treating or preventing the indication sought to be treated.

For viral gene therapy vectors, dosages will generally be from about 1 μ g to about 1 mg of nucleic acid per kg of body mass. For non-infective gene therapy vectors, dosages will generally be from about 1 μ g to about 100 mg of nucleic acid per kg of body mass. Antisense oligonucleotide dosages will generally be from about 1 μ g to about 100 mg of nucleic acid per kg of body mass.

Autoimmune Disorders

Autoimmune disorders whereby antibodies are produced against glycine transporters can be expected to be associated with disease states. For example, for the GlyT-2 transporters, such disorders can be expected to be associated with decreased muscle activity, for instance decreased muscle activity that clinically presents much like myasthenia gravis, or to be associated with decreased pain perception. See, for an example of a disease caused by autoantibodies to a molecule involved in neurotransmission (glutamic acid decarboxylase), Nathan et al., *J. Neurosci. Res.* 40: 134-137, 1995.

The presence of these antibodies can be measured by established immunological methods using protein sequences obtained from the nucleic acids described herein or the related glycine transporters reported elsewhere. See, for example, Kim et al., *Mol. Pharmacol.*, 45: 608-617, 1994 and Liu et al., *J. Biol. Chem.* 268: 22802-22808, 1992. Such immunological methods are described, for example, in Ausubel et al., *Short Protocols in Molecular Biology*, John Wiley & Sons, New York, 1992.

- 29 -

The following examples further illustrate the present invention, but of course, should not be construed as in any way limiting its scope.

Example 1A - GlyT-2 Cloning

The cDNA encoding human GlyT-2 was generated by Reverse-Transcription PCR (RT-PCR) in two steps. In the first step, a degenerate primer corresponding to the rat GlyT-2 nucleotide sequence from 2540 to 2521 (5'-GGRTCDATCATRTTYTTRTA) was used to prime cDNA synthesis from human spinal cord poly A mRNA (Clontech, Palo Alto, CA). The numbering recited herein for the rat sequence is according to the numbering reported in Liu et al., *J. Biol. Chem.* 268: 22802-22808, 1992. The following primer pairs were then used in PCR reactions:

Primer A1: 5'-CCNAARGARATGAAYAARCCNCC

(SEQ ID NO:37; based on NT 223-245 of rat sequence)

Primer A2: 5'-GCNGTGAAGTACACCACTTTNCC

(SEQ ID NO:38; based on NT 1490-1468 of rat sequence)

Primer B1: 5'-CCNAARGARATGAAYAARCCNCC

(SEQ ID NO:39; based on NT 223-245 of rat sequence; same primer as Primer A1)

Primer B2: 5'-GGCYTCNGGGTAARCCACRAANGC

(SEQ ID NO:40; based on NT 1872-1849 of rat sequence)

The designation "R" indicates that the oligonucleotide composition has a mixture of adenosine and guanosine at the indicated position; "N" is for mixed oligonucleotides with all four base combinations at the indicated position; "Y" is for mixtures of cytosine and thymidine; "K" is for mixtures of guanosine and thymidine; "D" is for mixtures of adenosine, guanosine and thymidine.

The fragments generated by the A1 + A2 primers and by the B1 + B2 primers were separately cloned into pCRscript (Stratagene, La Jolla, CA) or pCR2.1 (Invitrogen, San Diego, CA), and sequenced from the resulting clones using the AutoRead sequencing kit (Pharmacia, Piscataway, NJ). Comparison of these sequences to rat GlyT-2 using the Lipman-Pearson FASTA algorithm revealed a 89% identity, confirming that these sequences encoded human GlyT-2. The A1 + A2 primer pair produced clone phG2-1, which has the nucleic acid sequence of SEQ ID NO:5 as its insert. The B1 + B2 primer pair produced clone phG2-2, which has the nucleic acid sequence of SEQ ID NO:7 as its insert.

For the second step, cDNA was synthesized from human spinal cord or

- 30 -

crebellum mRNA (Clontech, Palo Alto, CA) using random hexamers (Promega, Madison, WI), and additional primers were designed based upon the sequence of clones phG2-1 and phG2-2 for PCR. The following primer pairs were used to amplify the 5' and 3' ends of the human GlyT-2 cDNA.

5 Primer C1: 5'-CGGTTCAATCTGTTTCCGATCAGACATG
(SEQ ID NO:41; based on NT 181-210 of rat sequence)

Primer C2: 5'-GCAGGCTCGCGCGTCCGCTG
(SEQ ID NO:42; based on NT 210-191 of human sequence)

10 Primer D1: 5'-CCCGTATGTCGTACTCGTGATCCTCCTCATCCG
(SEQ ID NO:43; based on NT 1284-1316 of human sequence)

Primer D2: 5'-CCNCCRTGNGTDATCATNGGRAANCCC
(SEQ ID NO:44; based on NT 2087-2061 of rat sequence)

Primer E1: 5'-CCCGTATGTCGTACTCGTGATCCTCCTCATCCG
(SEQ ID NO:43; based on NT 1284-1316 of human sequence)

15 Primer E2: 5'-CCATCCACACTACTGGAYYARCAYTGNGTNCC
(SEQ ID NO:45; based on NT 2624-2593 of rat sequence)

Primer F1: 5'-CAGATTTCTTCTCTTTATCTGCTGCATGG
(SEQ ID NO:46; based on NT 1417-1446 of human sequence)

Primer F2: 5'-GGRTCDATCATRTTYTTRTANCKYTCNCC
(SEQ ID NO:47; based on NT 2540-2512 of rat sequence)

20 Primer G1: 5'-CCTGCACCAACAGTGCCACAAGC
(SEQ ID NO:48; based on NT 1517-1539 of human sequence)

Primer G2: 5'-CCATCCACACTACTGGAYYARCAYTGNGTNCC
(SEQ ID NO:45; based on NT 2624-2593 of rat sequence)

25 Primer H1: 5'-CCAAGTACCTACGCACACACAAGCC
(SEQ ID NO:49; based on NT 1784-1808 of human sequence)

Primer H2: 5'-GGATTAATACGGGACCATCCACACTACT
(SEQ ID NO:50; based on NT 2638-2611 of rat sequence)

The C1 + C2 primer pair produced clones phG2-3-a and phG2-3-b which have the nucleic acid sequences of SEQ IDs 1 and 3 as their inserts, respectively. The D1 + D2 primer pair produced phG2-4-a and phGH2-4-b which have the nucleic acid sequences of SEQ IDs 10 and 12 as their inserts, respectively. The E1 + E2 primer pair produced a clone which is believed to encompass nucleotides 1317-2379. The F1 + F2 primer pair produced a clone which is believed to encompass nucleotides 1447-2298.

- 31 -

The G1 + G2 primer pair produced clone phG2-7-a, which has the nucleic acid sequence of SEQ ID NO:14 as its insert and clone phG2-7-b, which has the nucleic acid sequence of SEQ ID NO:16 as its insert. The H1 + H2 primer pair produced phG2-8-a and phGH2-8-b which have the nucleic acid sequences of SEQ IDs 22 and 24 as their inserts,
 5 respectively.

The PCR fragments were cloned into pCR2.1 (Invitrogen). Figure 1 shows the location of each of the cloned cDNAs in relation to the entire human GlyT-2 sequence. Clone phG2-3 and phG2-8b were obtained from human cerebellum mRNA while the rest were from spinal cord. The cDNA inserts were sequenced using the
 10 AutoRead sequencing kit (Pharmacia) and the ALFexpressTM automatic sequencing apparatus (Pharmacia). These sequences implied ten point variations in the amino acid sequence. Comparison of the human GlyT-2 DNA sequence of SEQ ID NO:18 to the rat GlyT-2 sequence revealed an 89% nucleic acid identity and a 94.4% amino acid identity using the FASTA algorithm.

15 Example 1A - Further GlyT-2 Cloning

The following primers were also employed:

Primer I1: 5'-AGCTCTGCGGGACTTGAGAG

(SEQ ID NO:51; based on NT 276-295 of human sequence)

Primer I2: 5'-GTACACCACTTTTCCTGAAGTCTTG

20 (SEQ ID NO:52; based on NT1245-1269 of human sequence)

Primer J1: 5'-AGCTCTGCGGGACTTGAGAG

(SEQ ID NO:51; based on NT 276-295 of human sequence)

Primer J1: 5'-CCTTGGTCTGCCACATTCTCAATGTTG

(SEQ ID NO:53; based on NT-1599-1625 of human sequence)

25 The I1 + I2 primer pair produced clones phG2-9-a, phG2-9-b and phG2-9-c which have the nucleic acid sequences of SEQ ID NOS:29, 31 and 33 as their inserts, respectively. The J1 + J2 primer pair produced clone phG2-10 which has the nucleic acid sequence of SEQ ID NO:35.

Example 2 - Full-length Clone

30 The human GlyT-2 cDNAs were then used to construct a full length human GlyT-2 coding sequence, which was cloned into the pcDNA3 vector (Invitrogen). The clone incorporated the nucleic acid sequence of SEQ ID NO:20 and was denoted pHGT2-a. The 5' end of the cDNA was constructed by inserting the 254 bp Hind III-
 Nar I fragment from clone phG2-3 into clone phG2-1, previously digested with Hind III

- 32 -

and Nar I. The 3' end of the cDNA was constructed by inserting the Hind III-Hinc II fragment from phG2-2 and the Hinc II-Xba I fragment from clone phG2-7 into the pcDNA3 vector previously digested with Hind III and Xba I. Lastly, the Hind III-Nru I fragment from the 5' end clone and the Nru I-Xba I fragment from the 3' end clone were
 5 cloned into the pcDNA3 vector (Invitrogen) digested with Hind III and Xba I. The pHGT2-a expression clone thus obtained contains the sequence of human GlyT-2 from 1 to 2397 under the control of the human cytomegalovirus (CMV) promoter. In this expression clone, nts 1-173 were derived from clone phG2-3; nts 174-823 were derived from clone phG2-1; nts 824-1599 were derived from clone phG2-2; and nts 1600-2397
 10 were derived from clone phG2-7 (see fig. 2).

Example 3A - Second Full-Length Clone

An expression clone containing the nucleic acid sequence of SEQ ID NO:18 is constructed from the expression clone containing SEQ ID NO:20 by site-directed mutagenesis to change NT 304 from G to A, NT 371 from T to C, NT 836 from A to T,
 15 NT 1116 from G to A, NT 1831 from G to A, NT 2382 from T to C, NT 2388 from A to G, NT 2391 from T to C and NT 2394 from A to G. The mutagenesis is conducted by the oligonucleotide-directed methodology described by Ausubel et al. *Current Protocols in Molecular Biology*, John Wiley and Sons, New York, 1995, pp.8.1.1-8.1.6.

Example 3B - Third Full-Length Clone

20 The human GlyT-2 cDNAs were used to construct another full-length GlyT-2 coding sequence, which was cloned into the pcDNA3 vector (Invitrogen). The clone, denoted pHGT2-b, incorporated the nucleic acid sequence of SEQ ID NO:28 and encoded SEQ ID NO:27. First, a 254 bp HindIII-NarI fragment from phG2-3a (SEQ ID NO:1) was inserted into clone phG2-2 (SEQ ID NO:7) which had previously been digested with
 25 HindIII-NarI, creating Intermediate 1. A 1.6 kb HindIII-HincII fragment from Intermediate 1 and an 800 bp HincII-XbaI fragment from clone phG2-7b were ligated into pcDNA that had been digested with HindIII-XbaI, creating Intermediate 2.

A NdeI-MscI fragment (1 kb) and a BsmI-NdeI fragment (6.9 kb, containing pcDNA3) from Intermediate 2 were ligated with a 434 bp MscI-BsmI fragment from
 30 phG2-1 (SEQ ID NO:5), creating Intermediate 3. A 3.8 kb BssHII fragment from Intermediate 3 was ligated with a 4.0 kb BssHII fragment of clone pHGT2-a (see Example 2), creating pHGT2-b. In pHGT2-b, nts 1-173 were derived from clone phG2-3a (SEQ ID NO:1), nts 174-523 and 962-1599 from clone phG2-2 (SEQ ID NO:7), nts 524-961 from clone phG2-1 (SEQ ID NO:5), and nts 1600-2397 from clone phG2-7b

- 33 -

(SEQ ID NO:16).

Example 4 - GlyT-2 Expression

The clones of examples 2 and 3B were transfected into QT-6 cells (from American Type Culture Collection, Accession No. ATCC CRL-1708) using the method described in Example 5. The glycine transport assay described in Example 6 was used to confirm that glycine transport activity was conferred to the cells by the transfection.

Example 5 - Transfection

This example sets forth methods and materials used for growing and transfecting QT-6 cells, which are avian fibroblasts derived from quail. Transfections with pHGT2-a have been conducted, as have transfections with GlyT-1 vectors, though these latter transfections were conducted at separate times.

QT-6 cells were obtained from American Type Culture Collection (Accession No. ATCC CRL-1708). Complete QT-6 medium for growing QT-6 was Medium 199 (Sigma Chemical Company, St. Louis, MO; hereinafter "Sigma") supplemented to be 10% tryptose phosphate; 5% fetal bovine serum (Sigma); 1% penicillin-streptomycin (Sigma); and 1% sterile dimethylsulfoxide (DMSO; Sigma). Other solutions required for growing or transfecting QT-6 cells included:

DNA/DEAE Mix: 450 μ l TBS, 450 μ l DEAE Dextran (Sigma), and 100 μ l of DNA (4 μ g) in TE, where the DNA included GlyT-1a, GlyT-1b, GlyT-1c, or GlyT-2 encoding DNA, in a suitable expression vector. The DNA used was as defined below.

PBS: Standard phosphate buffered saline, pH 7.4 including 1 mM CaCl_2 and 1 mM MgCl_2 sterilized through a 0.2 μ m filter.

TBS: One ml of Solution B, 10 ml of Solution A; brought to 100 ml with distilled H_2O ; filter-sterilized and stored at 4°C.

TE: 0.01 M Tris, 0.001 M EDTA, pH 8.0.

DEAE dextran: Sigma, #D-9885. A stock solution was prepared consisting of 0.1% (1 mg/ml) of the DEAE dextran in TBS. The stock solution was filter sterilized and frozen in 1 ml aliquots.

Chloroquine: Sigma, #C-6628. A stock solution was prepared consisting of 100 mM chloroquine in H_2O . The stock solution was filter-sterilized and stored in 0.5 ml aliquots, frozen.

- 34 -

Solution A (10X):

	NaCl	8.00 g
	KCl	0.38 g
	Na ₂ HPO ₄	0.20 g
5	Tris base	3.00 g

The solution was adjusted to pH 7.5 with HCl, brought to 100.0 ml with distilled H₂O, and filter-sterilized and stored at room temperature.

Solution B (100X):

	CaCl ₂ · 2H ₂ O	1.5 g
10	MgCl ₂ · 6H ₂ O	1.0 g

The solution was brought to 100 ml with distilled H₂O, and filter-sterilized: the solution was then stored at room temperature.

HBSS: 150 mM NaCl, 20 mM HEPES, 1 mM CaCl₂, 10 mM glucose, 5 mM KCl, 1 mM MgCl₂ · H₂O; adjusted with NaOH to pH 7.4.

15 Standard growth and passaging procedures used were as follows: Cells were grown in 225 ml flasks. For passaging, cells were washed twice with warm HBSS (5 ml each wash). Two ml of a 0.05% trypsin/EDTA solution was added, the culture was swirled, then the trypsin/EDTA solution was aspirated quickly. The culture was then incubated about 2 minutes (until cells lift off), then 10 ml of QT-6 media was added and

20 the cells are further dislodged by swirling the flask and tapping its bottom. The cells were removed and transferred to a 15 ml conical tube, centrifuged at 1000 xg for 10 minutes, and resuspended in 10 ml of QT-6 medium. A sample was removed for counting, the cells were then diluted further to a concentration of 1×10^5 cells/ml using QT-6 medium, and 65 ml of the culture was added per 225 ml flask of passaged cells.

25 Transfection was accomplished using cDNAs prepared as follows:

For human GlyT-2 expression, the pHGT2-a clone described above was used.

The human GlyT-1a (hGlyT-1a) clone contained the sequence of hGlyT-1a from nucleotide position 183 to 2108 cloned into the pRc/CMV vector (Invitrogen, San Diego, CA) as a Hind III-Xba I fragment as described in Kim et al., *Mol. Pharmacol.*, 45: 608-617, 1994. The first 17 nucleotides (corresponding to the first 6 amino acids) of the GlyT-1a sequence reported in this Kim et al. article is actually based on the rat sequence. To determine whether the sequence of human GlyT-1a is different in this region, the 5' region of hGlyT-1a from nucleotide 1 to 212 was obtained by rapid amplification of cDNA ends using the 5' RACE system supplied by Gibco BRL

- 35 -

(Gaithersburg, MD). Sequencing of this 5' region of GlyT-1a confirmed that the first 17 nucleotides of coding sequence are identical in human and rat GlyT-1a.

The human GlyT-1b (hGlyT-1b) clone contained the sequence of hGlyT-1b from nucleotide position 213 to 2274 cloned into the pRc/CMV vector as a Hind III - Xba I fragment as described in Kim et al., *supra*.

The human GlyT-1c (hGlyT-1c) clone contained the sequence of hGlyT-1c from nucleotide position 213 to 2336 cloned into the pRc/CMV vector (Invitrogen) as a Hind III - Xba I fragment as described in Kim et al., *supra*. The Hind III - Xba I fragment of hGlyT-1c from this clone was subcloned into the pRc/RSV vector.

Transfection experiments were performed with GlyT-1c in both the pRc/RSV and pRc/CMV expression vectors.

The following four day procedure for the transfections was used:

On day 1, QT-6 cells were plated at a density of 1×10^6 cells in 10 ml of complete QT-6 medium in 100 mm dishes.

On day 2, the medium was aspirated and the cells were washed with 10 ml of PBS followed by 10 ml of TBS. The TBS was aspirated, then 1 ml of the DEAE/DNA mix was added to the plate. The plate was swirled in the hood every 5 minutes. After 30 minutes, 8 ml of 80 μ M chloroquine in QT-6 medium was added and the culture was incubated for 2.5 hours at 37°C and 5% CO₂. The medium was then aspirated and the cells were washed two times with complete QT-6 medium, then 100 ml complete QT-6 medium was added and the cells were returned to the incubator.

On day 3, the cells were removed with trypsin/EDTA as described above, and plated into the wells of 96-well assay plates at approximately 2×10^5 cells/well.

On day 4, glycine transport was assayed as described in Example 6.

Example 6 - Glycine Uptake

This example illustrates a method for the measurement of glycine uptake by transfected cultured cells.

Transient GlyT-transfected cells or control cells grown in accordance with Example 5 were washed three times with HEPES buffered saline (HBS). The control cells were treated precisely as the GlyT-transfected cells except that the transfection procedure omitted any cDNA. The cells were incubated 10 minutes at 37°C, after which a solution was added containing 50 nM [³H] glycine (17.5 Ci/mmol) and either (a) no potential competitor, (b) 10 mM nonradioactive glycine or (c) a concentration of a prospective agent. A range of concentrations of the prospective agent was used to

- 36 -

generate data for calculating the concentration resulting in 50% of the effect (for example, the IC_{50} s, which are the concentrations of agent inhibiting glycine uptake by 50%). The cells were then incubated another 20 minutes at 37°C, after which the cells were washed three times with ice-cold HBS. Scintillant was added to the cells, the cells were shaken for 30 minutes, and the radioactivity in the cells was counted using a scintillation counter. Data were compared between the cells contacted or not contacted by a prospective agent, and, where relevant, between cells having GlyT-1 activity versus cells having GlyT-2 activity, depending on the assay being conducted.

Expression of glycine transporter activity in QT-6 cells transfected with the human GlyT-2 clone, pHGT2-a, is demonstrated in Figure 5, in which [3H] glycine uptake is shown for mock and pHGT2-a transfected cells. QT-6 cells transfected with pHGT2-a show significant increases in glycine transport as compared to mock transfected control cells. The results are presented as means \pm SEM of a representative experiment performed in triplicate. Substantially similar results were obtained with pHGT2-b.

The concentration dependence of glycine transport in pHGT2-a-transfected cells is shown in Figure 6: Substantially similar results were obtained with pHGT2-b. QT-6 cells transfected with the human GlyT-2 were incubated with 50 nM [3H] glycine and the indicated concentrations of unlabeled glycine for 20 minutes, and the cell-incorporated radioactivity was determined by scintillation counting. Data points represent means \pm SEM from an experiment performed in quadruplicate. The results indicated an IC_{50} of 40 μM .

Example 7 - Calcium Flux

This example illustrates a protocol for measuring calcium flux in cells.

The calcium flux measurement was generally performed in primary cell cultures, which were prepared using standard procedures and techniques that require sterile dissecting equipment, a microscope and defined medium. The protocol used was substantially as described by Lu et al., *Proc. Nat'l. Acad. Sci. USA*, 88: 6289-6292, 1991.

Example 8 - Binding to Strychnine-Sensitive Receptor

Binding of strychnine to strychnine-sensitive receptors was measured as described in White et al. *J. Neurochem.* 35: 503-512, 1989 and Becker et al., *J. Neurosci.* 6: 1358-1364, 1986, with minor modifications.

The nucleic acid (N.A.) r amino acid sequences referred to herein by SEQ ID NO: are as follows:

- 37 -

SEQUENCE LISTING

- (1) GENERAL INFORMATION
- (i) APPLICANT: Albert, Vivian
 - (ii) TITLE OF THE INVENTION: Human Glycine Transporter
 - (iii) NUMBER OF SEQUENCES: 53
 - (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Dechert Price & Rhoads
 - (B) STREET: 997 Lenox Drive, Building 3, Suite 210
 - (C) CITY: Lawrenceville
 - (D) STATE: NJ
 - (E) COUNTRY: USA
 - (F) ZIP: 08543
 - (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Diskette
 - (B) COMPUTER: IBM Compatible
 - (C) OPERATING SYSTEM: DOS
 - (D) SOFTWARE: FastSEQ for Windows Version 2.0
 - (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER:
 - (B) FILING DATE:
 - (C) CLASSIFICATION:
 - (vii) PRIOR APPLICATION DATA:
 - (A) APPLICATION NUMBER:
 - (B) FILING DATE:
 - (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Bloom, Allen
 - (B) REGISTRATION NUMBER: 29,135
 - (C) REFERENCE/DOCKET NUMBER: 317743-108WO
 - (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: 609-520-3214
 - (B) TELEFAX: 609-520-3259
 - (C) TELEX:
 - (2) INFORMATION FOR SEQ ID NO:1:
 - (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 190 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: linear
 - (xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```

ATGGATTGCA GTGCTCCCAA GGAAATGAAT AAAGTCCAG CCAACAGCCC GGAGGCGGCG      60
GCGGCGCAGG GCCACCCGGA TGGCCCATGC GCTCCAGGA CGAGCCCGGA GCAGGAGCTT      120
CCCGCGGCTG CCGCCCCGCC GCCGCCACGT GTGCCAGGT CCGCTTCCAC CGGCGCCCAA      180
ACTTTCCAGT                                     190

```

- (2) INFORMATION FOR SEQ ID NO:2:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 63 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```

Met Asp Cys Ser Ala Pro Lys Glu Met Asn Lys Leu Pro Ala Asn Ser
 1           5           10           15
Pro Glu Ala Ala Ala Gln Gly His Pro Asp Gly Pro Cys Ala Pro
          20           25           30
Arg Thr Ser Pro Glu Gln Glu Leu Pro Ala Ala Ala Pro Pro Pro
          35           40           45
Pro Arg Val Pro Arg Ser Ala Ser Thr Gly Ala Gln Thr Phe Gln
          50           55           60

```

- (2) INFORMATION FOR SEQ ID NO:3:
- (i) SEQUENCE CHARACTERISTICS:

- 38 -

- (A) LENGTH: 190 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

ATGGACTGCA	GTGCTCCCAA	GGAAATGAAT	AAACTGCCAG	CCAACAGCCC	GGAGGCGGCG	60
GCGGCGCAGG	GCCACCCGGA	TGGCCCATGC	GCTCCCAGGA	CGAGCCCGGA	GCAGGAGCTT	120
CCCGCGGCTG	CCGCCCCGCC	GCCGCCACGT	GTGCCCAGGT	CCGCTTCCAC	CGCGCCCCAA	180
ACTTTCCAGT						190

- (2) INFORMATION FOR SEQ ID NO:4:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 63 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met	Asp	Cys	Ser	Ala	Pro	Lys	Glu	Met	Asn	Lys	Leu	Pro	Ala	Asn	Ser
1				5				10					15		
Pro	Glu	Ala	Ala	Ala	Ala	Gln	Gly	His	Pro	Asp	Gly	Pro	Cys	Ala	Pro
			20				25					30			
Arg	Thr	Ser	Pro	Glu	Gln	Glu	Leu	Pro	Ala	Ala	Ala	Ala	Pro	Pro	Pro
		35				40					45				
Pro	Arg	Val	Pro	Arg	Ser	Ala	Ser	Thr	Gly	Ala	Gln	Thr	Phe	Gln	
	50				55						60				

- (2) INFORMATION FOR SEQ ID NO:5:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 1216 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

AGCCAACAGC	CCGGAGGCGG	CGGCGGCGCA	GGGCCACCCG	GATGGCCCAT	GCGCTCCCAG	60
GACGAGCCCG	GAGCAGGAGC	TTCCCGCGGC	TGCCGCCCCG	CCGCCGCCAC	GTGTGCCCAG	120
GTCCGCTTCC	ACCGGCGCCC	AAACTTTCCA	GTCAGCGGAC	GCGCGAGCCT	GCGAGGCTGA	180
GCGGCCAGGA	GTGGGGTCTT	GCAAACCTAG	TAGCCCGCGG	GCGCAGGCGG	CCTCTGCAGC	240
TCTGCGGGAC	TTGAGAGAGG	CGCAAGGCGC	GCAGGCCTCG	CCCCCTCCCG	GGAGCTCCGG	300
GCCCGGCAAC	GCGCTGCACT	GTAAGATCCC	TTTTCTGCGA	GGCCCGGAGG	GGGATGCGAA	360
CGTGAGTGTG	GGCAAGGGCA	CCCTGGAGCG	GAACAATACC	CCTGTTGTGG	GCTGGGTGAA	420
CATGAGCCAG	AGCACCGTGG	TGCTGGGCAC	GGATGGAATC	ACGTCCGTGC	TCCCGGGCAG	480
CGTGGCCACC	GTTGCCACCC	AGGAGGACGA	GCAAGGGGAT	GAGAATAAGG	CCCGAGGGAA	540
CTGGTCCAGC	AAACTGGACT	TCATCCTGTC	CATGGTGGGG	TACGCAGTGG	GGCTGGGCAA	600
TGTCTGGAGG	TTTCCCTACC	TGGCCTTCCA	GAACGGGGGA	GGTGCTTTCC	TCATCCCTTA	660
CCTGATGATG	CTGGCTCTGG	CTGGATTACC	CATCTTCTTC	TTGGAGGTGT	CGCTGGGCCA	720
GTTTGCCAGC	CAGGGACCAG	TGTCTGTGTG	GAAGGCCATC	CCAGCTCTAC	AAGGCTGTGG	780
CATCGCGATG	CTGATCATCT	CTGTCCTAAT	AGCCATATAC	TACAATGTGA	TTATTTGCTA	840
TACACTTTTC	TACCTGTTTG	CCTCCTTTGT	GTCTGTACTA	CCCTGGGGCT	CCTGCAACAA	900
CCCTTGGAAT	ACGCCAGAAT	GCAAAGATAA	AACCAAACCT	TTATTAGATT	CCTGTGTTAT	960
CAGTGACCAT	CCCAAATAC	AGATCAAGAA	CTCGACTTTC	TGCATGACCG	CTTATCCCAA	1020
CGTGACAATG	GTTAATTTCA	CCAGCCAGGC	CAATAAGACA	TTTGTCAGTG	GAAGTGAGAA	1080
GTAATTCAAG	TACTTTGTGC	TGAAGATTTT	TGCAGGGATT	GAATATCCTG	GCGAGATCGG	1140
GTGGCCACTA	GCTCTCTGCC	TCTTCCTGGC	TTGGGTGATT	GTGTATGCAT	CGTTGGCTAA	1200
AGGAATCAAG	ACTTCA					1216

- (2) INFORMATION FOR SEQ ID NO:6:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 405 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

- 39 -

Ala Asn Ser Pro Glu Ala Ala Ala Ala Gln Gly His Pro Asp Gly Pro
 1 5 10 15
 Cys Ala Pro Arg Thr Ser Pro Glu Gln Glu Leu Pro Ala Ala Ala Ala
 20 25 30
 Pro Pro Pro Pro Arg Val Pro Arg Ser Ala Ser Thr Gly Ala Gln Thr
 35 40 45
 Phe Gln Ser Ala Asp Ala Arg Ala Cys Glu Ala Glu Arg Pro Gly Val
 50 55 60
 Gly Ser Cys Lys Leu Ser Ser Pro Arg Ala Gln Ala Ala Ser Ala Ala
 65 70 75 80
 Leu Arg Asp Leu Arg Glu Ala Gln Gly Ala Gln Ala Ser Pro Pro Pro
 85 90 95
 Gly Ser Ser Gly Pro Gly Asn Ala Leu His Cys Lys Ile Pro Phe Leu
 100 105 110
 Arg Gly Pro Glu Gly Asp Ala Asn Val Ser Val Gly Lys Gly Thr Leu
 115 120 125
 Glu Arg Asn Asn Thr Pro Val Val Gly Trp Val Asn Met Ser Gln Ser
 130 135 140
 Thr Val Val Leu Gly Thr Asp Gly Ile Thr Ser Val Leu Pro Gly Ser
 145 150 155 160
 Val Ala Thr Val Ala Thr Gln Glu Asp Glu Gln Gly Asp Glu Asn Lys
 165 170 175
 Ala Arg Gly Asn Trp Ser Ser Lys Leu Asp Phe Ile Leu Ser Met Val
 180 185 190
 Gly Tyr Ala Val Gly Leu Gly Asn Val Trp Arg Phe Pro Tyr Leu Ala
 195 200 205
 Phe Gln Asn Gly Gly Gly Ala Phe Leu Ile Pro Tyr Leu Met Met Leu
 210 215 220
 Ala Leu Ala Gly Leu Pro Ile Phe Phe Leu Glu Val Ser Leu Gly Gln
 225 230 235 240
 Phe Ala Ser Gln Gly Pro Val Ser Val Trp Lys Ala Ile Pro Ala Leu
 245 250 255
 Gln Gly Cys Gly Ile Ala Met Leu Ile Ile Ser Val Leu Ile Ala Ile
 260 265 270
 Tyr Tyr Asn Val Ile Ile Cys Tyr Thr Leu Phe Tyr Leu Phe Ala Ser
 275 280 285
 Phe Val Ser Val Leu Pro Trp Gly Ser Cys Asn Asn Pro Trp Asn Thr
 290 295 300
 Pro Glu Cys Lys Asp Lys Thr Lys Leu Leu Leu Asp Ser Cys Val Ile
 305 310 315 320
 Ser Asp His Pro Lys Ile Gln Ile Lys Asn Ser Thr Phe Cys Met Thr
 325 330 335
 Ala Tyr Pro Asn Val Thr Met Val Asn Phe Thr Ser Gln Ala Asn Lys
 340 345 350
 Thr Phe Val Ser Gly Ser Glu Glu Tyr Phe Lys Tyr Phe Val Leu Lys
 355 360 365
 Ile Ser Ala Gly Ile Glu Tyr Pro Gly Glu Ile Gly Trp Pro Leu Ala
 370 375 380
 Leu Cys Leu Phe Leu Ala Trp Val Ile Val Tyr Ala Ser Leu Ala Lys
 385 390 395 400
 Gly Ile Lys Thr Ser
 405

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1597 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

AGCCAACAGC	CCGGAGGCGG	CGGCGGCGCA	GGGCCACCCG	GATGGCCCAT	GCGCTCCCAG	60
GACGAGCCCG	GAGCAGGAGC	TTCCCGCGGC	TGCCGCCCGG	CCGCCGCCAC	GTGTGCCAG	120
GTCCGCTTCC	ACCGGCGCCC	AAACTTTCCA	GTCAGCGGAC	GCGCGAGCCT	GCGAGGCTGA	180
GCGGCCAGGA	GTGGGGTCTT	GCAAATCAG	TAGCCCGCGG	GCGCAGGCGG	CCTCTGCAGC	240

- 40 -

TCTGCGGGAC	TTGAGAGAGG	CGCAAAGCGC	GCAGGCCTCG	CCCCCTCCCG	GGAGCTCCGG	300
GCCCGGCAAC	GCGCTGCACT	GTAAGATCCC	TTCTCTGCGA	GGCCCGGAGG	GGGATGCGAA	360
CGTGAGTGTG	GGCAAGGGCA	CCCTGGAGCG	GAACAATACC	CCTGTTGTGG	GCTGGGTGAA	420
CATGAGCCAG	AGCACCCTGG	TGCTGGGCAC	GGATGGAATC	ACGTCCGTGC	TCCCGGGCAG	480
CGTGGCCACC	GTTGCCACCC	AGGAGGACGA	GCAAGGGGAT	GAGAATAAGG	CCTGAGGGAA	540
CTGGTCCAGC	AAACTGGACT	TCATCCTGTC	CATGGTGGGG	TACGCAGTGG	GGCTGGGCAA	600
TGTCTGGAGG	TTTCCCTACC	TGGCCTTCCA	GAACGGGGGA	GGTGCTTCC	TCATCCCTTA	660
CCTGATGATG	CTGGCTCTGG	CTGGATTACC	CATCTTCTTC	TTGGAGGTGT	CGCTGGGCCA	720
GTTTGCCAGC	CAGGGACCAG	TGTCTGTGTG	GAAGGCCATC	CCAGCTCTAC	AAGGCTGTGG	780
CATCGCGATG	CTGATCAACT	CTGTCTTAAT	AGCCATATAC	TACAATGTGA	TTATTTGCTA	840
TACACTTTTC	TACCTGTTTG	CCTCCTTTGT	GTCTGTACTA	CCCTGGGGCT	CCTGCAACAA	900
CCCTTGGAAT	ACGCCAGAAT	GCAAAGATAA	AACCAAACTT	TTATTAGATT	CCTGTGTTAT	960
CAGTGACCAT	CCCAAATAC	AGATCAAGAA	CTCGACTTTC	TGCATGACCG	CTTATCCCAA	1020
CGTGACATG	GTAAATTTCA	CCAGCCAGGC	CAATAAGACA	TTTGTCAGTG	GAAGTGAGGA	1080
GTACTTCAAG	TACTTTGTGC	TGAAGATTTT	TGCAGGGATT	GAATATCTG	GCGAGATCAG	1140
GTGGCCACTA	GCTCTCTGCC	TCTTCCTGGC	TTGGGTCAAT	GTGTATGCAT	CGTTGGCTAA	1200
AGGAATCAAG	ACTTCAGGAA	AAGTGGTGTA	CTTCACGGCC	ACGTTCCCGT	ATGTCGTACT	1260
CGTGATCCTC	CTCATCCGAG	GAGTCACCCT	GCCTGGAGCT	GGAGCTGGGA	TCTGGTACTT	1320
CATCACACCC	AAGTGGGAGA	AACTCACGGA	TGCCACGGTG	TGGAAAGATG	CTGCCACTCA	1380
GATTTTCTTC	TCTTTATCTG	CTGCATGGGG	AGGCCTGATC	ACTCTCTCTT	CTTACAACAA	1440
ATTCCACAAC	AACTGCTACA	GGGACACTCT	AATTGTCACC	TGCACCAACA	GTGCCACAAG	1500
CATCTTTGCC	GGCTTCGTCA	TCTTCTCCGT	TATCGGCTTC	ATGGCCAATG	AACGCAAAGT	1560
CAACATTGAG	AATGTGGCAG	ACCAAGGGCC	AGGCATT			1597

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 177 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Ala	Asn	Ser	Pro	Glu	Ala	Ala	Ala	Ala	Gln	Gly	His	Pro	Asp	Gly	Pro
1				5					10					15	
Cys	Ala	Pro	Arg	Thr	Ser	Pro	Glu	Gln	Glu	Leu	Pro	Ala	Ala	Ala	Ala
			20					25					30		
Pro	Pro	Pro	Pro	Arg	Val	Pro	Arg	Ser	Ala	Ser	Thr	Gly	Ala	Gln	Thr
			35				40					45			
Phe	Gln	Ser	Ala	Asp	Ala	Arg	Ala	Cys	Glu	Ala	Glu	Arg	Pro	Gly	Val
	50				55					60					
Gly	Ser	Cys	Lys	Leu	Ser	Ser	Pro	Arg	Ala	Gln	Ala	Ala	Ser	Ala	Ala
65					70				75					80	
Leu	Arg	Asp	Leu	Arg	Glu	Ala	Gln	Ser	Ala	Gln	Ala	Ser	Pro	Pro	Pro
			85					90					95		
Gly	Ser	Ser	Gly	Pro	Gly	Asn	Ala	Leu	His	Cys	Lys	Ile	Pro	Ser	Leu
			100					105					110		
Arg	Gly	Pro	Glu	Gly	Asp	Ala	Asn	Val	Ser	Val	Gly	Lys	Gly	Thr	Leu
	115						120					125			
Glu	Arg	Asn	Asn	Thr	Pro	Val	Val	Gly	Trp	Val	Asn	Met	Ser	Gln	Ser
	130					135					140				
Thr	Val	Val	Leu	Gly	Thr	Asp	Gly	Ile	Thr	Ser	Val	Leu	Pro	Gly	Ser
145					150				155					160	
Val	Ala	Thr	Val	Ala	Thr	Gln	Glu	Asp	Glu	Gln	Gly	Asp	Glu	Asn	Lys
			165					170						175	
Ala															

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 354 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

- 41 -

Gly Asn Trp Ser Ser Lys Leu Asp Phe Ile Leu Ser Met Val Gly Tyr
 1 5 10 15
 Ala Val Gly Leu Gly Asn Val Trp Arg Phe Pro Tyr Leu Ala Phe Gln
 20 25 30
 Asn Gly Gly Gly Ala Phe Leu Ile Pro Tyr Leu Met Met Leu Ala Leu
 35 40 45
 Ala Gly Leu Pro Ile Phe Phe Leu Glu Val Ser Leu Gly Gln Phe Ala
 50 55 60
 Ser Gln Gly Pro Val Ser Val Trp Lys Ala Ile Pro Ala Leu Gln Gly
 65 70 75 80
 Cys Gly Ile Ala Met Leu Ile Asn Ser Val Leu Ile Ala Ile Tyr Tyr
 85 90 95
 Asn Val Ile Ile Cys Tyr Thr Leu Phe Tyr Leu Phe Ala Ser Phe Val
 100 105 110
 Ser Val Leu Pro Trp Gly Ser Cys Asn Asn Pro Trp Asn Thr Pro Glu
 115 120 125
 Cys Lys Asp Lys Thr Lys Leu Leu Leu Asp Ser Cys Val Ile Ser Asp
 130 135 140
 His Pro Lys Ile Gln Ile Lys Asn Ser Thr Phe Cys Met Thr Ala Tyr
 145 150 155 160
 Pro Asn Val Thr Met Val Asn Phe Thr Ser Gln Ala Asn Lys Thr Phe
 165 170 175
 Val Ser Gly Ser Glu Glu Tyr Phe Lys Tyr Phe Val Leu Lys Ile Ser
 180 185 190
 Ala Gly Ile Glu Tyr Pro Gly Glu Ile Arg Trp Pro Leu Ala Leu Cys
 195 200 205
 Leu Phe Leu Ala Trp Val Ile Val Tyr Ala Ser Leu Ala Lys Gly Ile
 210 215 220
 Lys Thr Ser Gly Lys Val Val Tyr Phe Thr Ala Thr Phe Pro Tyr Val
 225 230 235 240
 Val Leu Val Ile Leu Leu Ile Arg Gly Val Thr Leu Pro Gly Ala Gly
 245 250 255
 Ala Gly Ile Trp Tyr Phe Ile Thr Pro Lys Trp Glu Lys Leu Thr Asp
 260 265 270
 Ala Thr Val Trp Lys Asp Ala Ala Thr Gln Ile Phe Phe Ser Leu Ser
 275 280 285
 Ala Ala Trp Gly Gly Leu Ile Thr Leu Ser Ser Tyr Asn Lys Phe His
 290 295 300
 Asn Asn Cys Tyr Arg Asp Thr Leu Ile Val Thr Cys Thr Asn Ser Ala
 305 310 315 320
 Thr Ser Ile Phe Ala Gly Phe Val Ile Phe Ser Val Ile Gly Phe Met
 325 330 335
 Ala Asn Glu Arg Lys Val Asn Ile Glu Asn Val Ala Asp Gln Gly Pro
 340 345 350
 Gly Ile

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 533 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

AGGAGTCACC	CTGCCTGGAG	CTGGAGCTGG	GATCTGGTAC	TTCATCACAC	CCAACTGGGA	60
GAAACTCACG	GATGCCACGG	TGTGGAAAGA	TGCTGCCACT	CAGATTTTCT	TCTCTTTATC	120
TGCTGCATGG	GGAGGCCTGA	TCACTCTCTC	TTCTTACAAC	AAATTCCACA	ACAACTGCTA	180
CAGGGACACT	CTAATTGTCA	CCTGCACCAA	CAGTGCCACA	AGCATCTTTG	CCGGCTTCGT	240
CATCTTCTCC	GTTATCGGCT	TCATGGCCAA	TGAACGCAA	GTCAACATTG	AGAATGTGGC	300
AGACCAAGGG	CCAGGCATTG	CATTTGTGGT	TTACCCGGAA	GCCTTAACCA	GGCTGCCTCT	360
CTCTCCGTTT	TGGGCCATCA	TCTTTTCTCT	GATGCTCCTC	ACTCTGGGAC	TTGACACTAT	420
GTTTGCCACC	ATCGAGACCA	TAGTGACCTC	CATCTCAGAC	GAGTTTCCCA	AGTACCTACG	480
CACACACAAG	CCAGTGTTTA	CTCTGGGCTG	CTGCATTGTG	TTCTTCATCA	TGG	533

- 42 -

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 177 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

```

Gly Val Thr Leu Pro Gly Ala Gly Ala Gly Ile Trp Tyr Phe Ile Thr
 1           5           10           15
Pro Asn Trp Glu Lys Leu Thr Asp Ala Thr Val Trp Lys Asp Ala Ala
 20           25           30
Thr Gln Ile Phe Phe Ser Leu Ser Ala Ala Trp Gly Gly Leu Ile Thr
 35           40           45
Leu Ser Ser Tyr Asn Lys Phe His Asn Asn Cys Tyr Arg Asp Thr Leu
 50           55           60
Ile Val Thr Cys Thr Asn Ser Ala Thr Ser Ile Phe Ala Gly Phe Val
 65           70           75           80
Ile Phe Ser Val Ile Gly Phe Met Ala Asn Glu Arg Lys Val Asn Ile
 85           90           95
Glu Asn Val Ala Asp Gln Gly Pro Gly Ile Ala Phe Val Val Tyr Pro
100           105           110
Glu Ala Leu Thr Arg Leu Pro Leu Ser Pro Phe Trp Ala Ile Ile Phe
115           120           125
Phe Leu Met Leu Leu Thr Leu Gly Leu Asp Thr Met Phe Ala Thr Ile
130           135           140
Glu Thr Ile Val Thr Ser Ile Ser Asp Glu Phe Pro Lys Tyr Leu Arg
145           150           155           160
Thr His Lys Pro Val Phe Thr Leu Gly Cys Cys Ile Cys Phe Phe Ile
165           170           175
Met

```

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 533 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

```

AGGAGTCACC CTGCCTGGAG CTGGAGCTGG GATCTGGTAC TTCATCACAC CCAAGTGGGA      60
GAAACTCACG AATGCCACGG TGTGGAAAGA TGCTGCCACT CAGATTTTCT TCTCTTTATC      120
TGCTGCATGG GGAGGCCTGA TCACTCTCTC TTCTTACAAC AAATTCCACA ACAACTGCTA      180
CAGGGACACT CTAATTGTCA CCTGCACCAA CAGTGGCACA AGCATCTTTG CCGGCTTCGT      240
CATCTTCTCC GTTATCGGCT TCATGGCCAA TGAACGCAAA GTCAACATTG AGAATGTGGC      300
AGACCAAGGG CCAGGCATTG CATTGTGGT TTACCCGGAA GCCTTAACCA GGCTGCCTCT      360
CTCTCCGTTT TGGGCCATCA TCTTTTCTC GATGCTCCTC ACTCTGGAC TTGACACTAT      420
GTTTGCCACC ATCGAGACCA TAGTGACCTC CATCTCAGAC GAGTTTCCCA AGTACCTACG      480
CACACACAAG CCAGTGTTTA CTCTGGGCTG CTGCATTTGT TTCTTCATCA TGG          533

```

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 177 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

```

Gly Val Thr Leu Pro Gly Ala Gly Ala Gly Ile Trp Tyr Phe Ile Thr
 1           5           10           15
Pro Lys Trp Glu Lys Leu Thr Asn Ala Thr Val Trp Lys Asp Ala Ala
 20           25           30
Thr Gln Ile Phe Phe Ser Leu Ser Ala Ala Trp Gly Gly Leu Ile Thr
 35           40           45

```

- 43 -

Leu Ser Ser Tyr Asn Lys Phe His Asn Asn Cys Tyr Arg Asp Thr Leu
 50 55 60
 Ile Val Thr Cys Thr Asn Ser Ala Thr Ser Ile Phe Ala Gly Phe Val
 65 70 75 80
 Ile Phe Ser Val Ile Gly Phe Met Ala Asn Glu Arg Lys Val Asn Ile
 85 90 95
 Glu Asn Val Ala Asp Gln Gly Pro Gly Ile Ala Phe Val Val Tyr Pro
 100 105 110
 Glu Ala Leu Thr Arg Leu Pro Leu Ser Pro Phe Trp Ala Ile Ile Phe
 115 120 125
 Phe Leu Met Leu Leu Thr Leu Gly Leu Asp Thr Met Phe Ala Thr Ile
 130 135 140
 Glu Thr Ile Val Thr Ser Ile Ser Asp Glu Phe Pro Lys Tyr Leu Arg
 145 150 155 160
 Thr His Lys Pro Val Phe Thr Leu Gly Cys Cys Ile Cys Phe Phe Ile
 165 170 175
 Met

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 840 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

ATCTTTGCCG	GCTTCGTCAT	CTTCTCCGTT	ATCGGCTTCA	TGGCCAATGA	ACGCAAAGTC	60
AACATTGAGA	ATGTGGCAGA	CCAAGGGCCA	GGCATTGCAT	TTGTGGTTTA	CCCGGAAGCC	120
TTAACCAGGC	TGCCTCTCTC	TCCGTTCTGG	GCCATCATCT	TTTTCCTGAT	GCTCCTCACT	180
CTTGGAATTG	ACACTATGTT	TGCCACCATC	GAGACCATAG	TGACCTCCAT	CTCAGACGAG	240
TTTCCCAAGT	ACCTACGCAC	ACACAAGCCA	GTGTTTACTC	TGGGCTGCTG	CGTTTGTTC	300
TTCATCATGG	GTTTCCAAT	GATCACTCAG	GGTGAATTT	ACATGTTTCA	GCTTGTTGAC	360
ACCTATGCTG	CCTCCTATGC	CCTTGTCATC	ATTGCCATTT	TTGAGCTCGT	GGGGATCTCT	420
TATGTGTATG	GCTTGCAAAG	ATTCTGTGAA	GATATAGAGA	TGATGATTGG	ATTCCAGCCT	480
AACATCTTCT	GGAAAGTCTG	CTGGGCATTT	GTAACCCCAA	CCATTTTAAC	CTTTATCCTT	540
TGCTTCAGCT	TTACCAGTG	GGAGCCCATG	ACCTATGGCT	CTTACCGCTA	TCCTAACTGG	600
TCCATGGTGC	TGGATGGCT	AATGCTCGCC	TGTCCGTC	TCTGGATCCC	AATTATGTTT	660
GTGATAAAAA	TGCATCTGGC	CCCTGGAAGA	TTTATTGAGA	GGCTGAAGTT	GGTGTGCTCG	720
CCACAGCCGG	ACTGGGGCCC	ATTCTTAGCT	CAACACCGCG	GGGAGCGTTA	CAAGAACATG	780
ATCGACCCCT	TGGGAACCTC	TTCCTTGGGA	CTCAAACCTG	CAGTGAAGGA	TTTGGAACTG	840

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 280 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

Ile Phe Ala Gly Phe Val Ile Phe Ser Val Ile Gly Phe Met Ala Asn
 1 5 10 15
 Glu Arg Lys Val Asn Ile Glu Asn Val Ala Asp Gln Gly Pro Gly Ile
 20 25 30
 Ala Phe Val Val Tyr Pro Glu Ala Leu Thr Arg Leu Pro Leu Ser Pro
 35 40 45
 Phe Trp Ala Ile Ile Phe Phe Leu Met Leu Leu Thr Leu Gly Leu Asp
 50 55 60
 Thr Met Phe Ala Thr Ile Glu Thr Ile Val Thr Ser Ile Ser Asp Glu
 65 70 75 80
 Ph Pro Lys Tyr Leu Arg Thr His Lys Pro Val Phe Thr Leu Gly Cys
 85 90 95
 Cys Val Cys Phe Ph Ile Met Gly Phe Pro Met Ile Thr Gln Gly Gly
 100 105 110
 Ile Tyr Met Phe Gln Leu Val Asp Thr Tyr Ala Ala Ser Tyr Ala Leu

- 44 -

115	120	125
Val Ile Ala Ile Phe Glu Leu Val Gly Ile Ser Tyr Val Tyr Gly		
130	135	140
Leu Gln Arg Phe Cys Glu Asp Ile Glu Met Met Ile Gly Phe Gln Pro		
145	150	155
Asn Ile Phe Trp Lys Val Cys Trp Ala Phe Val Thr Pro Thr Ile Leu		
165	170	175
Thr Phe Ile Leu Cys Phe Ser Phe Tyr Gln Trp Glu Pro Met Thr Tyr		
180	185	190
Gly Ser Tyr Arg Tyr Pro Asn Trp Ser Met Val Leu Gly Trp Leu Met		
195	200	205
Leu Ala Cys Ser Val Ile Trp Ile Pro Ile Met Phe Val Ile Lys Met		
210	215	220
His Leu Ala Pro Gly Arg Phe Ile Glu Arg Leu Lys Leu Val Cys Ser		
225	230	235
Pro Gln Pro Asp Trp Gly Pro Phe Leu Ala Gln His Arg Gly Glu Arg		
245	250	255
Tyr Lys Asn Met Ile Asp Pro Leu Gly Thr Ser Ser Leu Gly Leu Lys		
260	265	270
Leu Pro Val Lys Asp Leu Glu Leu		
275	280	

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 840 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

ATCTTTGCCG	GCTTCGTCAT	CTTCTCCGTT	ATCGGCTTCA	TGGCCAATGA	ACGCAAAGTC	60
AACATTGAGA	ATGTGGCAGA	CCAAGGGCCA	GGCATTGCAT	TTGTGGTTTA	CCCGGAAGCC	120
TTAACCAGGC	TGCCTCTCTC	TCCGTTCTGG	GCCATCATCT	TTTTCTGAT	GCTCCTCACT	180
CTTGGACTTG	ACACTATGTT	TGCCACCATC	GAGACCATAG	TGACCTCCAT	CTCAGACGAG	240
TTTCCCAAGT	ACCTACGCAC	ACACAAGCCA	GTGTTTACTC	TGGGCTGCTG	CATTTGTTTC	300
TTCATCATGG	GTTTTCCAAT	GATCACTCAG	GGTGGAAATT	ACATGTTTCA	GCTTGTGGAC	360
ACCTATGCTG	CCTCCTATGC	CCTTGTCATC	ATTGCCATTT	TTGAGCTCGT	GGGGATCTCT	420
TATGTGTATG	GCTTGCAAAG	ATTCTGTGAA	GATATAGAGA	TGATGATTGG	ATTCCAGCCT	480
AACATCTTCT	GGAAAGTCTG	CTGGGCATTT	GTAACCCCAA	CCATTTTAAC	CTTTATCCTT	540
TGCTTCAGCT	TTTACCAGTG	GGAGCCCATG	ACCTATGGCT	CTTACCGCTA	TCCTAACTGG	600
TCCATGGTGC	TCGGATGGCT	AATGCTCGCC	TGTTCCGTC	TCTGGATCCC	AATTATGTTT	660
GTGATAAAAA	TGCATCTGGC	CCCTGGAAGA	TTTATTGAGA	GGCTGAAGTT	GGTGTGCTCG	720
CCACAGCCGG	ACTGGGGCCC	ATTCTTAGCT	CAACACCGCG	GGGAGCGTTA	CAAGAACATG	780
ATCGACCCCT	TGGGAACCTC	TTCCTTGGGA	CTCAAACCTG	CAGTGAAGGA	TTTGGAACTG	840

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 280 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

Ile Phe Ala Gly Phe Val Ile Phe Ser Val Ile Gly Phe Met Ala Asn	
1	5
Glu Arg Lys Val Asn Ile Glu Asn Val Ala Asp Gln Gly Pro Gly Ile	
20	25
Ala Phe Val Val Tyr Pro Glu Ala Leu Thr Arg Leu Pro Leu Ser Pro	
35	40
Phe Trp Ala Ile Ile Phe Phe Leu Met Leu Leu Thr Leu Gly Leu Asp	
50	55
Thr Met Phe Ala Thr Ile Glu Thr Ile Val Thr Ser Ile Ser Asp Glu	
65	70
Phe Pro Lys Tyr Leu Arg Thr His Lys Pro Val Phe Thr Leu Gly Cys	
85	90
	95

- 45 -

Cys Ile Cys Phe Phe Ile Met Gly Phe Pro Met Ile Thr Gln Gly Gly
 100 105 110
 Ile Tyr Met Phe Gln Leu Val Asp Thr Tyr Ala Ala Ser Tyr Ala Leu
 115 120 125
 Val Ile Ile Ala Ile Phe Glu Leu Val Gly Ile Ser Tyr Val Tyr Gly
 130 135 140
 Leu Gln Arg Phe Cys Glu Asp Ile Glu Met Met Ile Gly Phe Gln Pro
 145 150 155 160
 Asn Ile Phe Trp Lys Val Cys Trp Ala Phe Val Thr Pro Thr Ile Leu
 165 170 175
 Thr Phe Ile Leu Cys Phe Ser Phe Tyr Gln Trp Glu Pro Met Thr Tyr
 180 185 190
 Gly Ser Tyr Arg Tyr Pro Asn Trp Ser Met Val Leu Gly Trp Leu Met
 195 200 205
 Leu Ala Cys Ser Val Ile Trp Ile Pro Ile Met Phe Val Ile Lys Met
 210 215 220
 His Leu Ala Pro Gly Arg Phe Ile Glu Arg Leu Lys Leu Val Cys Ser
 225 230 235 240
 Pro Gln Pro Asp Trp Gly Pro Phe Leu Ala Gln His Arg Gly Glu Arg
 245 250 255
 Tyr Lys Asn Met Ile Asp Pro Leu Gly Thr Ser Ser Leu Gly Leu Lys
 260 265 270
 Leu Pro Val Lys Asp Leu Glu Leu
 275 280

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2397 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

ATGGATTGCA	GTGCTCCCAA	GGAAATGAAT	AAACTGCCAG	CCAACAGCCC	GGAGGCGGCG	60
GCGGCGCAGG	GCCACCCGGA	TGGCCCATGC	GCTCCCAGGA	CGAGCCCGGA	GCAGGAGCTT	120
CCCGCGGCTG	CCGCCCCGCC	GCCGCCACGT	GTGCCCAGGT	CCGCTTCCAC	CGGCGCCCAA	180
ACTTTCCAGT	CAGCGGACGC	GCGAGCCTGC	GAGGCTGAGC	GGCCAGGAGT	GGGGTCTTGC	240
AAACTCAGTA	GCCCGCGGGC	GCAGGCGGCC	TCTGCAGCTC	TGCGGGACTT	GAGAGAGGCG	300
CAAAGCGCGC	AGGCCTCGCC	CCCTCCCGGG	AGCTCCGGGC	CCGGCAACGC	GCTGCACTGT	360
AAGATCCCTT	CTCTGCGAGG	CCCGGAGGGG	GATGCGAACG	TGAGTGTGGG	CAAGGGCACC	420
CTGGAGCGGA	ACAATACCCC	TGTTGTGGGC	TGGGTGAACA	TGAGCCAGAG	CACCGTGGTG	480
CTGGGCACGG	ATGGAATCAC	GTCCGTGCTC	CCGGGCAGCG	TGGCCACCGT	TGCCACCCAG	540
GAGGACGAGC	AAGGGGATGA	GAATAAGGCC	CGAGGGAAC	GGTCCAGCAA	ACTGGACTTC	600
ATCCTGTCCA	TGGTGGGGTA	CGCAGTGGGG	CTGGGCAATG	TCTGGAGGTT	TCCCTACCTG	660
GCCTTCCAGA	ACGGGGGAGG	TGCTTTCCTC	ATCCCTTACC	TGATGATGCT	GGCTCTGGCT	720
GGATTACCCA	TCTTCTTCTT	GGAGGTGTCT	CTGGGCCAGT	TTGCCAGCCA	GGGACCAGTG	780
TCTGTGTGGA	AGGCCATCCC	AGCTCTACAA	GGCTGTGGCA	TCGCGATGCT	GATCATCTCT	840
GTCCTAATAG	CCATATACTA	CAATGTGATT	ATTTGCTATA	CACTTTCTTA	CCTGTTTGCC	900
TCCTTTGTGT	CTGTACTACC	CTGGGGCTCC	TGCAACAACC	CTTGAATAC	GCCAGAATGC	960
AAAGATAAAA	CCAAACTTTT	ATTAGATTCC	TGTGTTATCA	GTGACCATCC	CAAAATACAG	1020
ATCAAGAACT	CGACTTCTTG	CATGACCGCT	TATCCCAACG	TGACAATGGT	TAATTTTACC	1080
AGCCAGGCCA	ATAAGACATT	TGTCAAGTGA	AGTGAAGAGT	ACTTCAAGTA	CTTTGTGCTG	1140
AAGATTTCTG	CAGGGATTGA	ATATCCTGGC	GAGATCAGGT	GGCCACTAGC	TCTCTGCCTC	1200
TTCTTGCTTT	GGGTCAATGT	GTATGCATCG	TTGGCTAAAG	GAATCAAGAC	TTCAGGAAAA	1260
GTGGTGTACT	TCACGGCCAC	GTTCCCGTAT	GTGCTACTCG	TGATCCTCCT	CATCCGAGGA	1320
GTCACCCCTG	CTGGAGCTGG	AGCTGGGATC	TGGTACTTCA	TCACACCCAA	GTGGGAGAAA	1380
CTCACGGATG	CCACGGTGTG	GAAAGATGCT	GCCACTCAGA	TTTTCTTCTC	TTTATCTGTG	1440
GCATGGGGAG	GCCTGATCAC	TCTCTCTTCT	TACAACAAAT	TCCACAACAA	CTGCTACAGG	1500
GACACTCTAA	TTGTACCTTG	CACCAACAGT	GCCACAAGCA	TCTTTGCCGG	CTTCGTCATC	1560
TTCTCCGTTA	TCGGCTTCAT	GGCCAATGAA	CGCAAAGTCA	ACATTGAGAA	TGTGGCAGAC	1620
CAAGGGCCAG	GCATTGCATT	TGTGGTTTAC	CCGGAAGCCT	TAACCAGGCT	GCCTCTCTCT	1680
CCGTTCTGGG	CCATCATCTT	TTTCTGTAGT	CTCCTCACTG	TTGGACTTGA	CACTATGTTT	1740
GCCACCATCG	AGACCATAGT	GACCTCCATC	TCAGACGAGT	TTCCCAAGTA	CCTACGCACA	1800
CACAAGCCAG	TGTTTACTCT	GGGCTGCTGC	ATTTGTTTCT	TCATCATGGG	TTTTCCAATG	1860
ATCACTCAGG	GTGGAATTTA	CATGTTTCAG	CTTGTGGACA	CCTATGCTGC	CTCCTATGCC	1920

- 46 -

```

CTTGTCATCA TTGCCATTTT TGAGCTCGTG GGGATCTCTT ATGTGTATGG CTGCAAAGA 1980
TTCTGTGAAG ATATAGAGAT GATGATTGGA TTCCAGCCTA ACATCTTCTG GAAAGTCTGC 2040
TGGGCATTTG TAACCCCAAC CATTTTAACC TTTATCCTTT GCTTCAGCTT TTACCAAGTG 2100
GAGCCCATGA CCTATGGCTC TTACCGCTAT CCTAACTGGT CCATGGTGCT CGGATGGCTA 2160
ATGCTCGCCT GTTCCGTCAT CTGGATCCCA ATTATGTTTG TGATAAAAAT GCATCTGGCC 2220
CCTGGAAGAT TTATTGAGAG GCTGAAGTTG GTGTGCTCGC CACAGCCGGA CTGGGGCCCA 2280
TTCTTAGCTC AACACCGCGG GGAGCGTTAC AAGAACATGA TCGACCCCTT GGGAACCTCT 2340
TCCTTGGGAC TCAAAGTACC AGTGAAGGAT TTGGAAGTGG GCACTCAGTG CTAGTCC 2397

```

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 797 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

```

Met Asp Cys Ser Ala Pro Lys Glu Met Asn Lys Leu Pro Ala Asn Ser
 1      5      10      15
Pro Glu Ala Ala Ala Ala Gln Gly His Pro Asp Gly Pro Cys Ala Pro
 20      25      30
Arg Thr Ser Pro Glu Gln Glu Leu Pro Ala Ala Ala Pro Pro Pro
 35      40      45
Pro Arg Val Pro Arg Ser Ala Ser Thr Gly Ala Gln Thr Phe Gln Ser
 50      55      60
Ala Asp Ala Arg Ala Cys Glu Ala Glu Arg Pro Gly Val Gly Ser Cys
 65      70      75      80
Lys Leu Ser Ser Pro Arg Ala Gln Ala Ala Ser Ala Ala Leu Arg Asp
 85      90      95
Leu Arg Glu Ala Gln Ser Ala Gln Ala Ser Pro Pro Pro Gly Ser Ser
100      105      110
Gly Pro Gly Asn Ala Leu His Cys Lys Ile Pro Ser Leu Arg Gly Pro
115      120      125
Glu Gly Asp Ala Asn Val Ser Val Gly Lys Gly Thr Leu Glu Arg Asn
130      135      140
Asn Thr Pro Val Val Gly Trp Val Asn Met Ser Gln Ser Thr Val Val
145      150      155      160
Leu Gly Thr Asp Gly Ile Thr Ser Val Leu Pro Gly Ser Val Ala Thr
165      170      175
Val Ala Thr Gln Glu Asp Glu Gln Gly Asp Glu Asn Lys Ala Arg Gly
180      185      190
Asn Trp Ser Ser Lys Leu Asp Phe Ile Leu Ser Met Val Gly Tyr Ala
195      200      205
Val Gly Leu Gly Asn Val Trp Arg Phe Pro Tyr Leu Ala Phe Gln Asn
210      215      220
Gly Gly Gly Ala Phe Leu Ile Pro Tyr Leu Met Met Leu Ala Leu Ala
225      230      235
Gly Leu Pro Ile Phe Phe Leu Glu Val Ser Leu Gly Gln Phe Ala Ser
245      250      255
Gln Gly Pro Val Ser Val Trp Lys Ala Ile Pro Ala Leu Gln Gly Cys
260      265      270
Gly Ile Ala Met Leu Ile Ile Ser Val Leu Ile Ala Ile Tyr Tyr Asn
275      280      285
Val Ile Ile Cys Tyr Thr Leu Phe Tyr Leu Phe Ala Ser Phe Val Ser
290      295      300
Val Leu Pro Trp Gly Ser Cys Asn Asn Pro Trp Asn Thr Pro Glu Cys
305      310      315
Lys Asp Lys Thr Lys Leu Leu Leu Asp Ser Cys Val Ile Ser Asp His
325      330      335
Pro Lys Ile Gln Ile Lys Asn Ser Thr Phe Cys Met Thr Ala Tyr Pro
340      345      350
Asn Val Thr Met Val Asn Phe Thr Ser Gln Ala Asn Lys Thr Phe Val
355      360      365
Ser Gly Ser Glu Glu Tyr Phe Lys Tyr Phe Val Leu Lys Ile Ser Ala
370      375      380

```

- 47 -

Gly Ile Glu Tyr Pro Gly Glu Ile Arg Trp Pro Leu Ala Leu Cys Leu
 385 390 395 400
 Phe Leu Ala Trp Val Ile Val Tyr Ala Ser Leu Ala Lys Gly Ile Lys
 405 410 415
 Thr Ser Gly Lys Val Val Tyr Phe Thr Ala Thr Phe Pro Tyr Val Val
 420 425 430
 Leu Val Ile Leu Leu Ile Arg Gly Val Thr Leu Pro Gly Ala Gly Ala
 435 440 445
 Gly Ile Trp Tyr Phe Ile Thr Pro Lys Trp Glu Lys Leu Thr Asp Ala
 450 455 460
 Thr Val Trp Lys Asp Ala Ala Thr Gln Ile Phe Phe Ser Leu Ser Ala
 465 470 475 480
 Ala Trp Gly Gly Leu Ile Thr Leu Ser Ser Tyr Asn Lys Phe His Asn
 485 490 495
 Asn Cys Tyr Arg Asp Thr Leu Ile Val Thr Cys Thr Asn Ser Ala Thr
 500 505 510
 Ser Ile Phe Ala Gly Phe Val Ile Phe Ser Val Ile Gly Phe Met Ala
 515 520 525
 Asn Glu Arg Lys Val Asn Ile Glu Asn Val Ala Asp Gln Gly Pro Gly
 530 535 540
 Ile Ala Phe Val Val Tyr Pro Glu Ala Leu Thr Arg Leu Pro Leu Ser
 545 550 555 560
 Pro Phe Trp Ala Ile Ile Phe Phe Leu Met Leu Leu Thr Leu Gly Leu
 565 570 575
 Asp Thr Met Phe Ala Thr Ile Glu Thr Ile Val Thr Ser Ile Ser Asp
 580 585 590
 Glu Phe Pro Lys Tyr Leu Arg Thr His Lys Pro Val Phe Thr Leu Gly
 595 600 605
 Cys Cys Ile Cys Phe Phe Ile Met Gly Phe Pro Met Ile Thr Gln Gly
 610 615 620
 Gly Ile Tyr Met Phe Gln Leu Val Asp Thr Tyr Ala Ala Ser Tyr Ala
 625 630 635 640
 Leu Val Ile Ile Ala Ile Phe Glu Leu Val Gly Ile Ser Tyr Val Tyr
 645 650 655
 Gly Leu Gln Arg Phe Cys Glu Asp Ile Glu Met Met Ile Gly Phe Gln
 660 665 670
 Pro Asn Ile Phe Trp Lys Val Cys Trp Ala Phe Val Thr Pro Thr Ile
 675 680 685
 Leu Thr Phe Ile Leu Cys Phe Ser Phe Tyr Gln Trp Glu Pro Met Thr
 690 695 700
 Tyr Gly Ser Tyr Arg Tyr Pro Asn Trp Ser Met Val Leu Gly Trp Leu
 705 710 715 720
 Met Leu Ala Cys Ser Val Ile Trp Ile Pro Ile Met Phe Val Ile Lys
 725 730 735
 Met His Leu Ala Pro Gly Arg Phe Ile Glu Arg Leu Lys Leu Val Cys
 740 745 750
 Ser Pro Gln Pro Asp Trp Gly Pro Phe Leu Ala Gln His Arg Gly Glu
 755 760 765
 Arg Tyr Lys Asn Met Ile Asp Pro Leu Gly Thr Ser Ser Leu Gly Leu
 770 775 780
 Lys Leu Pro Val Lys Asp Leu Glu Leu Gly Thr Gln Cys
 785 790 795

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2397 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

ATGGATTGCA	GTGCTCCCAA	GGAAATGAAT	AAACTGCCAG	CCAACAGCCC	GGAGGCGGCG	60
GCGGCGCAGG	GCCACCCGGA	TGGCCCATGC	GCTCCCAGGA	CGAGCCCGGA	GCAGGAGCTT	120
CCCGCGGCTG	CCGCCCCGCC	GCCGCCACGT	GTGCCCAGGT	CCGCTTCCAC	CGGCGCCCAA	180
ACTTTCCAGT	CAGCGGACGC	GCGAGCCTGC	GAGGCTGAGC	GGCCAGGAGT	GGGGTCTTGC	240

- 48 -

AAACTCAGTA	GCCCCGCGGC	GCAGGCGGCC	TCTGCAGCTC	TGCGGGACTT	GAGAGAGGCG	300
CAAGGCGCGC	AGGCTCGCC	CCCTCCCGGG	AGCTCCGGGC	CCGGCAACGC	GCTGCACTGT	360
AAGATCCCTT	TTCTGCGAGG	CCCGGAGGGG	GATGCCGAACG	TGAGTGTGGG	CAAGGGCACC	420
CTGGAGCGGA	ACAATACCCC	TGTTGTGGGC	TGGGTGAACA	TGAGCCAGAG	CACCGTGGTG	480
CTGGGCACGG	ATGGAATCAC	GTCCGTGCTC	CCGGGCAGCG	TGGCCACCGT	TGCCACCCAG	540
GAGGACGAGC	AAGGGGATGA	GAATAAGGCC	CGAGGGAAC	GGTCCAGCAA	ACTGGACTTC	600
ATCCTGTCCA	TGGTGGGGTA	CGCAGTGGGG	CTGGGCAATG	TCTGGAGGTT	TCCCTACCTG	660
GCCTTCCAGA	ACGGGGGAGG	TGCTTTCCTC	ATCCCTTACC	TGATGATGCT	GGCTCTGGCT	720
GGATTACCCA	TCTTCTTCTT	GGAGGTGTCG	CTGGGCCAGT	TTGCCAGCCA	GGGACCAGTG	780
TCTGTGTGGA	AGGCCATCCC	AGCTCTACAA	GGCTGTGGCA	TCGCGATGCT	GATCAACTCT	840
GTCCCTAATAG	CCATATACTA	CAATGTGATT	ATTTGCTATA	CACCTTTCTA	CCTGTTTGCC	900
TCCTTTGTGT	CTGTACTACC	CTGGGGCTCC	TGCAACAACC	CTTGGAATAC	GCCAGAATGC	960
AAAGATAAAA	CCAAACTTTT	ATTAGATTCC	TGTGTTATCA	GTGACCATCC	CAAAATACAG	1020
ATCAAGAACT	CGACTTTCTG	CATGACCGCT	TATCCCAACG	TGACAATGGT	TAATTTACCC	1080
AGCCAGGCCA	ATAAGACATT	TGTCAGTGGA	AGTGAGGAGT	ACTTCAAGTA	CTTTGTGCTG	1140
AAGATTTCTG	CAGGGATTGA	ATATCCTGGC	GAGATCAGGT	GGCCACTAGC	TCTCTGCCTC	1200
TTCTGGGCTT	GGGTCAATTG	GTATGCATCG	TTGGCTAAAG	GAATCAAGAC	TTCAAGGAAA	1260
GTGGTGTACT	TCACGGCCAC	GTTCCCGTAT	GTCTGACTCG	TGATCCTCCT	CATCCGAGGA	1320
GTCACCCTGC	CTGGAGCTGG	AGCTGGGATC	TGGTACTTCA	TCACACCCAA	GTGGGAGAAA	1380
CTCACGGATG	CCACGGTGTG	GAAAGATGCT	GCCACTCAGA	TTTTCTTCTC	TTTATCTGCT	1440
GCATGGGGAG	GCCTGATCAC	TCTCTCTTCT	TACAACAAAT	TCCACAACAA	CTGCTACAGG	1500
GACACTCTAA	TTGTACCTG	CACCAACAGT	GCCACAAGCA	TCTTTGCCGG	CTTCGTATC	1560
TTCTCCGTTA	TCGGCTTCAT	GGCCAATGAA	CGCAAAGTCA	ACATTGAGAA	TGTGGCAGAC	1620
CAAGGGCCAG	GCATTGCATT	TGTGGTTTAC	CCGGAAGCCT	TAACCAGGCT	GCCTCTCTCT	1680
CCGTTCTGGG	CCATCATCTT	TTTCCTGATG	CTCCTCACTC	TTGGACTTGA	CACTATGTTT	1740
GCCACCATCG	AGACCATAGT	GACCTCCATC	TCAGACGAGT	TTCCCAAGTA	CCTACGCACA	1800
CACAAGCCAG	TGTTTACTCT	GGGCTGCTGC	GTTTGTCTCT	TCATCATGGG	TTTTCCAATG	1860
ATCACTCAGG	GTGGAATTTA	CATGTTTCAG	CTTGTGGACA	CCTATGCTGC	CTCCTATGCC	1920
CTTGTGCATCA	TTGCCATTTT	TGAGCTCGTG	GGGATCTCTT	ATGTGTATGG	CTTGCAAAGA	1980
TTCTGTGAAG	ATATAGAGAT	GATGATTGGA	TTCCAGCCTA	ACATCTTCTG	GAAAGTCTGC	2040
TGGGCATTTG	TAACCCCAAC	CATTTTAACC	TTTATCCTTT	GCTTCAGCTT	TTACCAAGTG	2100
GAGCCCATGA	CCTATGGCTC	TTACCGCTAT	CCTAATGGT	CCATGGTGCT	CGGATGGCTA	2160
ATGCTGCGCT	GTTCCGTCAT	CTGGATCCCA	ATTATGTTTG	TGATAAAAAT	GCATCTGGCC	2220
CCTGGAAGAT	TTATTGAGAG	GCTGAAGTTG	GTGTGCTCGC	CACAGCCGGA	CTGGGGCCCA	2280
TTCTTAGCTC	AACACCGCGG	GGAGCGTTAC	AAGAACATGA	TCGACCCCTT	GGGAACCTCT	2340
TCCTTGGGAC	TCAAACCTGCC	AGTGAAGGAT	TTGGAACCTG	GTACTCAATG	TTAATCC	2397

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 797 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

Met	Asp	Cys	Ser	Ala	Pro	Lys	Glu	Met	Asn	Lys	Leu	Pro	Ala	Asn	Ser
1				5				10						15	
Pro	Glu	Ala	Ala	Ala	Ala	Gln	Gly	His	Pro	Asp	Gly	Pro	Cys	Ala	Pro
			20					25					30		
Arg	Thr	Ser	Pro	Glu	Gln	Glu	Leu	Pro	Ala	Ala	Ala	Ala	Pro	Pro	Pro
		35					40					45			
Pro	Arg	Val	Pro	Arg	Ser	Ala	Ser	Thr	Gly	Ala	Gln	Thr	Phe	Gln	Ser
		50				55					60				
Ala	Asp	Ala	Arg	Ala	Cys	Glu	Ala	Glu	Arg	Pro	Gly	Val	Gly	Ser	Cys
65					70					75				80	
Lys	Leu	Ser	Ser	Pro	Arg	Ala	Gln	Ala	Ala	Ser	Ala	Ala	Leu	Arg	Asp
			85					90						95	
Leu	Arg	Glu	Ala	Gln	Gly	Ala	Gln	Ala	Ser	Pro	Pro	Pro	Gly	Ser	Ser
			100					105					110		
Gly	Pro	Gly	Asn	Ala	Leu	His	Cys	Lys	Ile	Pro	Phe	Leu	Arg	Gly	Pro
		115					120					125			
Glu	Gly	Asp	Ala	Asn	Val	Ser	Val	Gly	Lys	Gly	Thr	Leu	Glu	Arg	Asn
		130				135					140				
Asn	Thr	Pro	Val	Val	Gly	Trp	Val	Asn	Met	Ser	Gln	Ser	Thr	Val	Val
145					150					155					160

- 49 -

Leu Gly Thr Asp Gly Ile Thr Ser Val Leu Pro Gly Ser Val Ala Thr
 165 170 175
 Val Ala Thr Gln Glu Asp Glu Gln Gly Asp Glu Asn Lys Ala Arg Gly
 180 185 190
 Asn Trp Ser Ser Lys Leu Asp Phe Ile Leu Ser Met Val Gly Tyr Ala
 195 200 205
 Val Gly Leu Gly Asn Val Trp Arg Phe Pro Tyr Leu Ala Phe Gln Asn
 210 215 220
 Gly Gly Gly Ala Phe Leu Ile Pro Tyr Leu Met Met Leu Ala Leu Ala
 225 230 235 240
 Gly Leu Pro Ile Phe Phe Leu Glu Val Ser Leu Gly Gln Phe Ala Ser
 245 250 255
 Gln Gly Pro Val Ser Val Trp Lys Ala Ile Pro Ala Leu Gln Gly Cys
 260 265 270
 Gly Ile Ala Met Leu Ile Asn Ser Val Leu Ile Ala Ile Tyr Tyr Asn
 275 280 285
 Val Ile Ile Cys Tyr Thr Leu Phe Tyr Leu Phe Ala Ser Phe Val Ser
 290 295 300
 Val Leu Pro Trp Gly Ser Cys Asn Asn Pro Trp Asn Thr Pro Glu Cys
 305 310 315 320
 Lys Asp Lys Thr Lys Leu Leu Leu Asp Ser Cys Val Ile Ser Asp His
 325 330 335
 Pro Lys Ile Gln Ile Lys Asn Ser Thr Phe Cys Met Thr Ala Tyr Pro
 340 345 350
 Asn Val Thr Met Val Asn Phe Thr Ser Gln Ala Asn Lys Thr Phe Val
 355 360 365
 Ser Gly Ser Glu Glu Tyr Phe Lys Tyr Phe Val Leu Lys Ile Ser Ala
 370 375 380
 Gly Ile Glu Tyr Pro Gly Glu Ile Arg Trp Pro Leu Ala Leu Cys Leu
 385 390 395 400
 Phe Leu Ala Trp Val Ile Val Tyr Ala Ser Leu Ala Lys Gly Ile Lys
 405 410 415
 Thr Ser Gly Lys Val Val Tyr Phe Thr Ala Thr Phe Pro Tyr Val Val
 420 425 430
 Leu Val Ile Leu Leu Ile Arg Gly Val Thr Leu Pro Gly Ala Gly Ala
 435 440 445
 Gly Ile Trp Tyr Phe Ile Thr Pro Lys Trp Glu Lys Leu Thr Asp Ala
 450 455 460
 Thr Val Trp Lys Asp Ala Ala Thr Gln Ile Phe Phe Ser Leu Ser Ala
 465 470 475 480
 Ala Trp Gly Gly Leu Ile Thr Leu Ser Ser Tyr Asn Lys Phe His Asn
 485 490 495
 Asn Cys Tyr Arg Asp Thr Leu Ile Val Thr Cys Thr Asn Ser Ala Thr
 500 505 510
 Ser Ile Phe Ala Gly Phe Val Ile Phe Ser Val Ile Gly Phe Met Ala
 515 520 525
 Asn Glu Arg Lys Val Asn Ile Glu Asn Val Ala Asp Gln Gly Pro Gly
 530 535 540
 Ile Ala Phe Val Val Tyr Pro Glu Ala Leu Thr Arg Leu Pro Leu Ser
 545 550 555 560
 Pro Phe Trp Ala Ile Ile Phe Phe Leu Met Leu Leu Thr Leu Gly Leu
 565 570 575
 Asp Thr Met Phe Ala Thr Ile Glu Thr Ile Val Thr Ser Ile Ser Asp
 580 585 590
 Glu Phe Pro Lys Tyr Leu Arg Thr His Lys Pro Val Phe Thr Leu Gly
 595 600 605
 Cys Cys Val Cys Phe Phe Ile Met Gly Phe Pro Met Ile Thr Gln Gly
 610 615 620
 Gly Ile Tyr Met Phe Gln Leu Val Asp Thr Tyr Ala Ala Ser Tyr Ala
 625 630 635 640
 Leu Val Ile Ile Ala Ile Phe Glu Leu Val Gly Ile Ser Tyr Val Tyr
 645 650 655
 Gly Leu Gln Arg Phe Cys Glu Asp Ile Glu Met Met Ile Gly Phe Gln
 660 665 670
 Pro Asn Ile Phe Trp Lys Val Cys Trp Ala Phe Val Thr Pro Thr Ile

- 50 -

```

      675      680      685
Leu Thr Phe Ile Leu Cys Phe Ser Phe Tyr Gln Trp Glu Pro Met Thr
 690      695      700
Tyr Gly Ser Tyr Arg Tyr Pro Asn Trp Ser Met Val Leu Gly Trp Leu
 705      710      715      720
Met Leu Ala Cys Ser Val Ile Trp Ile Pro Ile Met Phe Val Ile Lys
      725      730      735
Met His Leu Ala Pro Gly Arg Phe Ile Glu Arg Leu Lys Leu Val Cys
      740      745      750
Ser Pro Gln Pro Asp Trp Gly Pro Phe Leu Ala Gln His Arg Gly Glu
      755      760      765
Arg Tyr Lys Asn Met Ile Asp Pro Leu Gly Thr Ser Ser Leu Gly Leu
      770      775      780
Lys Leu Pro Val Lys Asp Leu Glu Leu Gly Thr Gln Cys
 785      790      795

```

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 589 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

```

AGTGTCTTACT CTGGGCTGCT ACATTTGTTT CTTTCATCATG GGTTTTCCAA TGATCACTCA      60
GGGTGGAATT TACATGTTTC AGCTTGTGGA CACCTATGCT GCCTCCTATG CCCTTGTTCAT      120
CATTGCCATT TTTGAGCTCG TGGGGATCTC TTATGTGTAT GGCTTGCAAA GATTCTGTGA      180
AGATATAGAG ATGATGATTG GATTCCAGCC TAACATCTTC TGGAAAGTCT GCTGGGCATT      240
TGTAACCCCA ACCATTTTAA CCTTTATCCT TTGCTTCAGC TTTTACCAGT GGGAGCCCAT      300
GACCTATGGC TCTTACCGCT ATCCTAACTG GTCCATGGTG CTCGGATGGC TAATGCTCGC      360
CTGTTCCGTC ATCTGGATCC CAATTATGTT TGTGGTAAAA ATGCATCTGG CCCCTGGAAG      420
ATTTATTGAG AGGCTGAAGT TGGTGTGCTC GCCACAGCCG GACTGGGGCC CATTCTTAGC      480
TCAACACCGC GGGGAGCGTT ACAAGAACAT GATCGACCCC TTGGGAACCT CTTCTTGGG      540
ACTCAAACTG CCAGTGAAGG ATTTGGAAGT GGGCACTCAG TGCTAGTCC      589

```

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 194 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

```

Val Phe Thr Leu Gly Cys Tyr Ile Cys Phe Phe Ile Met Gly Phe Pro
 1      5      10      15
Met Ile Thr Gln Gly Gly Ile Tyr Met Phe Gln Leu Val Asp Thr Tyr
      20      25      30
Ala Ala Ser Tyr Ala Leu Val Ile Ile Ala Ile Phe Glu Leu Val Gly
      35      40      45
Ile Ser Tyr Val Tyr Gly Leu Gln Arg Phe Cys Glu Asp Ile Glu Met
      50      55      60
Met Ile Gly Phe Gln Pro Asn Ile Phe Trp Lys Val Cys Trp Ala Phe
      65      70      75      80
Val Thr Pro Thr Ile Leu Thr Phe Ile Leu Cys Phe Ser Phe Tyr Gln
      85      90      95
Trp Glu Pro Met Thr Tyr Gly Ser Tyr Arg Tyr Pro Asn Trp Ser Met
      100      105      110
Val Leu Gly Trp Leu Met Leu Ala Cys Ser Val Ile Trp Ile Pro Ile
      115      120      125
Met Phe Val Val Lys Met His Leu Ala Pro Gly Arg Phe Ile Glu Arg
      130      135      140
Leu Lys Leu Val Cys Ser Pro Gln Pro Asp Trp Gly Pro Phe Leu Ala
      145      150      155      160
Gln His Arg Gly Glu Arg Tyr Lys Asn Met Ile Asp Pro Leu Gly Thr
      165      170      175

```

- 51 -

Ser Ser Leu Gly Leu Lys Leu Pro Val Lys Asp Leu Glu Leu Gly Thr
 180 185 190
 Gln Cys

(2) INFORMATION FOR SEQ ID NO:24:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 589 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

AGTGTCTTACT	CTGGGCTGCT	GCATTTGTTT	CTTCATCATG	GGTTTTCCAA	TGATCACTCA	60
GGGTGGAATT	TACATGTTTC	AGCTTGTTGA	CACCTATGCT	GCCTCCTATG	CCCTTGTCAT	120
CATTGCCATT	TTTGAGCTCG	TGGGGATCTC	TTATGTGTAT	GGCTTGCAAA	GATTCTGTGA	180
AGATATAGAG	ATGATGATTG	GATTCCAGCC	TAACATCTTC	TGGAAAGTCT	GCTGGGCATT	240
TGTAACCCCA	ACCATTTTAA	CCTTTATCCT	TTGCTTCAGC	TTTTACCAGT	GGGAACCCAT	300
GACCTATGGC	TCTTACCGCT	ATCCTAACTG	GTCCATGGTG	CTCGGATGGC	TAATGCTCGC	360
CTGTTCCGTC	ATCTGGATCC	CAATTATGTC	TGTGATAAAA	ATGCATCTGG	CCCCTGGAAG	420
ATTTATTGAG	AGGCTGAAGT	TGGTGTGCTC	GCCACAGCCG	GACTGGGGCC	CATTCTTAGC	480
TCAACACCGC	GGGGAGCGTT	ACAAGAACAT	GATCGACCCC	TTGGGAACCT	CTTCCTTGGG	540
ACTCAAACCTG	CCAGTGAAGG	ATTTGGAAC	GGGCACTCAG	TGCTAGTCC		589

(2) INFORMATION FOR SEQ ID NO:25:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 194 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

Val	Phe	Thr	Leu	Gly	Cys	Cys	Ile	Cys	Phe	Phe	Ile	Met	Gly	Phe	Pro
1				5					10					15	
Met	Ile	Thr	Gln	Gly	Gly	Ile	Tyr	Met	Phe	Gln	Leu	Val	Asp	Thr	Tyr
			20					25					30		
Ala	Ala	Ser	Tyr	Ala	Leu	Val	Ile	Ile	Ala	Ile	Phe	Glu	Leu	Val	Gly
			35				40					45			
Ile	Ser	Tyr	Val	Tyr	Gly	Leu	Gln	Arg	Phe	Cys	Glu	Asp	Ile	Glu	Met
			50			55					60				
Met	Ile	Gly	Phe	Gln	Pro	Asn	Ile	Phe	Trp	Lys	Val	Cys	Trp	Ala	Phe
65					70					75				80	
Val	Thr	Pro	Thr	Ile	Leu	Thr	Phe	Ile	Leu	Cys	Phe	Ser	Phe	Tyr	Gln
				85					90					95	
Trp	Glu	Pro	Met	Thr	Tyr	Gly	Ser	Tyr	Arg	Tyr	Pro	Asn	Trp	Ser	Met
			100				105						110		
Val	Leu	Gly	Trp	Leu	Met	Leu	Ala	Cys	Ser	Val	Ile	Trp	Ile	Pro	Ile
			115				120					125			
Met	Ser	Val	Ile	Lys	Met	His	Leu	Ala	Pro	Gly	Arg	Phe	Ile	Glu	Arg
			130			135					140				
Leu	Lys	Leu	Val	Cys	Ser	Pro	Gln	Pro	Asp	Trp	Gly	Pro	Phe	Leu	Ala
145				150						155				160	
Gln	His	Arg	Gly	Glu	Arg	Tyr	Lys	Asn	Met	Ile	Asp	Pro	Leu	Gly	Thr
			165				170						175		
Ser	Ser	Leu	Gly	Leu	Lys	Leu	Pro	Val	Lys	Asp	Leu	Glu	Leu	Gly	Thr
			180				185						190		
Gln	Cys														

(2) INFORMATION FOR SEQ ID NO:26:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2397 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

- 52 -

(ix) FEATURE:

(A) NAME/KEY: Coding Sequence

(B) LOCATION: 1...2391

(D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

ATG GAT TGC AGT GCT CCC AAG GAA ATG AAT AAA CTG CCA GCC AAC AGC	48
Met Asp Cys Ser Ala Pro Lys Glu Met Asn Lys Leu Pro Ala Asn Ser	
1 5 10 15	
CCG GAG GCG GCG GCG GCG CAG GGC CAC CCG GAT GGC CCA TGC GCT CCC	96
Pro Glu Ala Ala Ala Ala Gln Gly His Pro Asp Gly Pro Cys Ala Pro	
20 25 30	
AGG ACG AGC CCG GAG CAG GAG CTT CCC GCG GCT GCC GCC CCG CCG CCG	144
Arg Thr Ser Pro Glu Gln Glu Leu Pro Ala Ala Ala Pro Pro Pro	
35 40 45	
CCA CGT GTG CCC AGG TCC GCT TCC ACC GGC GCC CAA ACT TTC CAG TCA	192
Pro Arg Val Pro Arg Ser Ser Thr Gly Ala Gln Thr Phe Gln Ser	
50 55 60	
GCG GAC GCG CGA GCC TGC GAG GCT GAG CGG CCA GGA GTG GGG TCT TGC	240
Ala Asp Ala Arg Ala Cys Glu Ala Glu Arg Pro Gly Val Gly Ser Cys	
65 70 75 80	
AAA CTC AGT AGC CCG CGG GCG CAG GCG GCC TCT GCA GCT CTG CGG GAC	288
Lys Leu Ser Ser Pro Arg Ala Gln Ala Ala Ser Ala Ala Leu Arg Asp	
85 90 95	
TTG AGA GAG GCG CAA GGC GCG CAG GCC TCG CCC CCT CCC GGG AGC TCC	336
Leu Arg Glu Ala Gln Gly Ala Gln Ala Ser Pro Pro Pro Gly Ser Ser	
100 105 110	
GGG CCC GGC AAC GCG CTG CAC TGT AAG ATC CCT TCT CTG CGA GGC CCG	384
Gly Pro Gly Asn Ala Leu His Cys Lys Ile Pro Ser Leu Arg Gly Pro	
115 120 125	
GAG GGG GAT GCG AAC GTG AGT GTG GGC AAG GGC ACC CTG GAG CGG AAC	432
Glu Gly Asp Ala Asn Val Ser Val Gly Lys Gly Thr Leu Glu Arg Asn	
130 135 140	
AAT ACC CCT GTT GTG GGC TGG GTG AAT ATG AGC CAG AGC ACC GTG GTG	480
Asn Thr Pro Val Val Gly Trp Val Asn Met Ser Gln Ser Thr Val Val	
145 150 155 160	
CTG GGC ACG GAT GGA ATC ACG TCC GTG CTC CCG GGC AGC GTG GCC ACC	528
Leu Gly Thr Asp Gly Ile Thr Ser Val Leu Pro Gly Ser Val Ala Thr	
165 170 175	
GTT GCC ACC CAG GAG GAC GAG CAA GGG GAT GAG AAT AAG GCC CGA GGG	576
Val Ala Thr Gln Glu Asp Glu Gln Gly Asp Glu Asn Lys Ala Arg Gly	
180 185 190	
AAC TGG TCC AGC AAA CTG GAC TTC ATC CTG TCC ATG GTG GGG TAC GCA	624
Asn Trp Ser Ser Lys Leu Asp Phe Ile Leu Ser Met Val Gly Tyr Ala	
195 200 205	
GTG GGG CTG GGC AAT GTC TGG AGG TTT CCC TAC CTG GCC TTC CAG AAC	672
Val Gly Leu Gly Asn Val Trp Arg Phe Pro Tyr Leu Ala Phe Gln Asn	
210 215 220	
GGG GGA GGT GCT TTC CTC ATC CCT TAC CTG ATG ATG CTG GCT CTG GCT	720
Gly Gly Gly Ala Phe Leu Ile Pro Tyr Leu Met Met Leu Ala Leu Ala	
225 230 235 240	

- 53 -

GGA TTA CCC ATC TTC TTC TTG GAG GTG TCG CTG GGC CAG TTT GCC AGC Gly Leu Pro Ile Phe Phe Leu Glu Val Ser Leu Gly Gln Phe Ala Ser	768
245 250 255	
CAG GGA CCA GTG TCT GTG TGG AAG GCC ATC CCA GCT CTA CAA GGC TCT Gln Gly Pro Val Ser Val Trp Lys Ala Ile Pro Ala Leu Gln Gly Cys	816
260 265 270	
GGC ATC GCG ATG CTG ATC ATC TCT GTC CTA ATA GCC ATA TAC TAC AAT Gly Ile Ala Met Leu Ile Ile Ser Val Leu Ile Ala Ile Tyr Tyr Asn	864
275 280 285	
GTG ATT ATT TGC TAT ACA CTT TTC TAC CTG TTT GCC TCC TTT GTG TCT Val Ile Ile Cys Tyr Thr Leu Phe Tyr Leu Phe Ala Ser Phe Val Ser	912
290 295 300	
GTA CTA CCC TGG GGC TCC TGC AAC AAC CCT TGG AAT ACG CCA GAA TGC Val Leu Pro Trp Gly Ser Cys Asn Asn Pro Trp Asn Thr Pro Glu Cys	960
305 310 315 320	
AAA GAT AAA ACC AAA CTT TTA TTA GAT TCC TGT GTT ATC AGT GAC CAT Lys Asp Lys Thr Lys Leu Leu Leu Asp Ser Cys Val Ile Ser Asp His	1008
325 330 335	
CCC AAA ATA CAG ATC AAG AAC TCG ACT TTC TGC ATG ACC GCT TAT CCC Pro Lys Ile Gln Ile Lys Asn Ser Thr Phe Cys Met Thr Ala Tyr Pro	1056
340 345 350	
AAC GTG ACA ATG GTT AAT TTC ACC AGC CAG GCC AAT AAG ACA TTT GTC Asn Val Thr Met Val Asn Phe Thr Ser Gln Ala Asn Lys Thr Phe Val	1104
355 360 365	
AGT GGA AGT GAA GAG TAC TTC AAG TAC TTT GTG CTG AAG ATT TCT GCA Ser Gly Ser Glu Glu Tyr Phe Lys Tyr Phe Val Leu Lys Ile Ser Ala	1152
370 375 380	
GGG ATT GAA TAT CCT GGC GAG ATC AGG TGG CCA CTA GCT CTC TGC CTC Gly Ile Glu Tyr Pro Gly Glu Ile Arg Trp Pro Leu Ala Leu Cys Leu	1200
385 390 395 400	
TTC CTG GCT TGG GTC ATT GTG TAT GCA TCG TTG GCT AAA GGA ATC AAG Phe Leu Ala Trp Val Ile Val Tyr Ala Ser Leu Ala Lys Gly Ile Lys	1248
405 410 415	
ACT TCA GGA AAA GTG GTG TAC TTC ACG GCC ACG TTC CCG TAT GTC GTA Thr Ser Gly Lys Val Val Tyr Phe Thr Ala Thr Phe Pro Tyr Val Val	1296
420 425 430	
CTC GTG ATC CTC CTC ATC CGA GGA GTC ACC CTG CCT GGA GCT GGA GCT Leu Val Ile Leu Leu Ile Arg Gly Val Thr Leu Pro Gly Ala Gly Ala	1344
435 440 445	
GGG ATC TGG TAC TTC ATC ACA CCC AAG TGG GAG AAA CTC ACG GAT GCC Gly Ile Trp Tyr Phe Ile Thr Pro Lys Trp Glu Lys Leu Thr Asp Ala	1392
450 455 460	
ACG GTG TGG AAA GAT GCT GCC ACT CAG ATT TTC TTC TCT TTA TCT GCT Thr Val Trp Lys Asp Ala Ala Thr Gln Ile Phe Phe Ser Leu Ser Ala	1440
465 470 475 480	
GCA TGG GGA GGC CTG ATC ACT CTC TCT TCT TAC AAC AAA TTC CAC AAC Ala Trp Gly Gly Leu Ile Thr Leu Ser Ser Tyr Asn Lys Phe His Asn	1488
485 490 495	
AAC TGC TAC AGG GAC ACT CTA ATT GTC ACC TGC ACC AAC AGT GCC ACA	1536

- 54 -

Asn	Cys	Tyr	Arg	Asp	Thr	Leu	Ile	Val	Thr	Cys	Thr	Asn	Ser	Ala	Thr	
			500					505					510			
AGC	ATC	TTT	GCC	GGC	TTC	GTC	ATC	TTC	TCC	GTT	ATC	GGC	TTC	ATG	GCC	1584
Ser	Ile	Phe	Ala	Gly	Phe	Val	Ile	Phe	Ser	Val	Ile	Gly	Phe	Met	Ala	
		515					520					525				
AAT	GAA	CGC	AAA	GTC	AAC	ATT	GAG	AAT	GTG	GCA	GAC	CAA	GGG	CCA	GGC	1632
Asn	Glu	Arg	Lys	Val	Asn	Ile	Glu	Asn	Val	Ala	Asp	Gln	Gly	Pro	Gly	
	530					535					540					
ATT	GCA	TTT	GTG	GTT	TAC	CCG	GAA	GCC	TTA	ACC	AGG	CTG	CCT	CTC	TCT	1680
Ile	Ala	Phe	Val	Val	Tyr	Pro	Glu	Ala	Leu	Thr	Arg	Leu	Pro	Leu	Ser	
545					550					555					560	
CCG	TTC	TGG	GCC	ATC	ATC	TTT	TTC	CTG	ATG	CTC	CTC	ACT	CTT	GGA	CTT	1728
Pro	Phe	Trp	Ala	Ile	Ile	Phe	Phe	Leu	Met	Leu	Leu	Thr	Leu	Gly	Leu	
				565					570					575		
GAC	ACT	ATG	TTT	GCC	ACC	ATC	GAG	ACC	ATA	GTG	ACC	TCC	ATC	TCA	GAC	1776
Asp	Thr	Met	Phe	Ala	Thr	Ile	Glu	Thr	Ile	Val	Thr	Ser	Ile	Ser	Asp	
			580					585					590			
GAG	TTT	CCC	AAG	TAC	CTA	CGC	ACA	CAC	AAG	CCA	GTG	TTT	ACT	CTG	GGC	1824
Glu	Phe	Pro	Lys	Tyr	Leu	Arg	Thr	His	Lys	Pro	Val	Phe	Thr	Leu	Gly	
		595					600					605				
TGC	TGC	ATT	TGT	TTC	TTC	ATC	ATG	GGT	TTT	CCA	ATG	ATC	ACT	CAG	GGT	1872
Cys	Cys	Ile	Cys	Phe	Phe	Ile	Met	Gly	Phe	Pro	Met	Ile	Thr	Gln	Gly	
	610					615					620					
GGA	ATT	TAC	ATG	TTT	CAG	CTT	GTG	GAC	ACC	TAT	GCT	GCC	TCC	TAT	GCC	1920
Gly	Ile	Tyr	Met	Phe	Gln	Leu	Val	Asp	Thr	Tyr	Ala	Ala	Ser	Tyr	Ala	
625					630					635					640	
CTT	GTC	ATC	ATT	GCC	ATT	TTT	GAG	CTC	GTG	GGG	ATC	TCT	TAT	GTG	TAT	1968
Leu	Val	Ile	Ile	Ala	Ile	Phe	Glu	Leu	Val	Gly	Ile	Ser	Tyr	Val	Tyr	
				645					650					655		
GGC	TTG	CAA	AGA	TTC	TGT	GAA	GAT	ATA	GAG	ATG	ATG	ATT	GGA	TTC	CAG	2016
Gly	Leu	Gln	Arg	Phe	Cys	Glu	Asp	Ile	Glu	Met	Met	Ile	Gly	Phe	Gln	
			660				665						670			
CCT	AAC	ATC	TTC	TGG	AAA	GTC	TGC	TGG	GCA	TTT	GTA	ACC	CCA	ACC	ATT	2064
Pro	Asn	Ile	Phe	Trp	Lys	Val	Cys	Trp	Ala	Phe	Val	Thr	Pro	Thr	Ile	
		675				680						685				
TTA	ACC	TTT	ATC	CTT	TGC	TTC	AGC	TTT	TAC	CAG	TGG	GAG	CCC	ATG	ACC	2112
Leu	Thr	Phe	Ile	Leu	Cys	Phe	Ser	Phe	Tyr	Gln	Trp	Glu	Pro	Met	Thr	
	690					695					700					
TAT	GGC	TCT	TAC	CGC	TAT	CCT	AAC	TGG	TCC	ATG	GTG	CTC	GGA	TGG	CTA	2160
Tyr	Gly	Ser	Tyr	Arg	Tyr	Pro	Asn	Trp	Ser	Met	Val	Leu	Gly	Trp	Leu	
705					710					715					720	
ATG	CTC	GCC	TGT	TCC	GTC	ATC	TGG	ATC	CCA	ATT	ATG	TTT	GTG	ATA	AAA	2208
Met	Leu	Ala	Cys	Ser	Val	Ile	Trp	Ile	Pro	Ile	Met	Phe	Val	Ile	Lys	
				725					730					735		
ATG	CAT	CTG	GCC	CCT	GGA	AGA	TTT	ATT	GAG	AGG	CTG	AAG	TTG	GTG	TGC	2256
M t	His	Leu	Ala	Pro	Gly	Arg	Phe	Ile	Glu	Arg	Leu	Lys	Leu	Val	Cys	
			740					745					750			
TCG	CCA	CAG	CCG	GAC	TGG	GGC	CCA	TTC	TTA	GCT	CAA	CAC	CGC	GGG	GAG	2304
Ser	Pro	Gln	Pro	Asp	Trp	Gly	Pro	Phe	Leu	Ala	Gln	His	Arg	Gly	Glu	

- 55 -

755	760	765	
CGT TAC AAG AAC ATG ATC GAC CCC TTG GGA ACC TCT TCC TTG GGA CTC			2352
Arg Tyr Lys Asn Met Ile Asp Pro Leu Gly Thr Ser Ser Leu Gly Leu			
770	775	780	
AAA CTG CCA GTG AAG GAT TTG GAA CTG GGC ACT CAG TGC TAGTCC			2397
Lys Leu Pro Val Lys Asp Leu Glu Leu Gly Thr Gln Cys			
785	790	795	

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 797 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

Met	Asp	Cys	Ser	Ala	Pro	Lys	Glu	Met	Asn	Lys	Leu	Pro	Ala	Asn	Ser
1				5				10						15	
Pro	Glu	Ala	Ala	Ala	Ala	Gln	Gly	His	Pro	Asp	Gly	Pro	Cys	Ala	Pro
			20				25						30		
Arg	Thr	Ser	Pro	Glu	Gln	Glu	Leu	Pro	Ala	Ala	Ala	Ala	Pro	Pro	Pro
		35				40						45			
Pro	Arg	Val	Pro	Arg	Ser	Ala	Ser	Thr	Gly	Ala	Gln	Thr	Phe	Gln	Ser
	50					55					60				
Ala	Asp	Ala	Arg	Ala	Cys	Glu	Ala	Glu	Arg	Pro	Gly	Val	Gly	Ser	Cys
	65				70					75				80	
Lys	Leu	Ser	Ser	Pro	Arg	Ala	Gln	Ala	Ala	Ser	Ala	Ala	Leu	Arg	Asp
			85					90						95	
Leu	Arg	Glu	Ala	Gln	Gly	Ala	Gln	Ala	Ser	Pro	Pro	Pro	Gly	Ser	Ser
			100					105					110		
Gly	Pro	Gly	Asn	Ala	Leu	His	Cys	Lys	Ile	Pro	Ser	Leu	Arg	Gly	Pro
		115					120					125			
Glu	Gly	Asp	Ala	Asn	Val	Ser	Val	Gly	Lys	Gly	Thr	Leu	Glu	Arg	Asn
	130					135					140				
Asn	Thr	Pro	Val	Val	Gly	Trp	Val	Asn	Met	Ser	Gln	Ser	Thr	Val	Val
	145				150					155				160	
Leu	Gly	Thr	Asp	Gly	Ile	Thr	Ser	Val	Leu	Pro	Gly	Ser	Val	Ala	Thr
			165					170						175	
Val	Ala	Thr	Gln	Glu	Asp	Glu	Gln	Gly	Asp	Glu	Asn	Lys	Ala	Arg	Gly
			180					185					190		
Asn	Trp	Ser	Ser	Lys	Leu	Asp	Phe	Ile	Leu	Ser	Met	Val	Gly	Tyr	Ala
		195				200						205			
Val	Gly	Leu	Gly	Asn	Val	Trp	Arg	Phe	Pro	Tyr	Leu	Ala	Phe	Gln	Asn
	210					215					220				
Gly	Gly	Gly	Ala	Phe	Leu	Ile	Pro	Tyr	Leu	Met	Met	Leu	Ala	Leu	Ala
	225				230					235				240	
Gly	Leu	Pro	Ile	Phe	Phe	Leu	Glu	Val	Ser	Leu	Gly	Gln	Phe	Ala	Ser
			245						250					255	
Gln	Gly	Pro	Val	Ser	Val	Trp	Lys	Ala	Ile	Pro	Ala	Leu	Gln	Gly	Cys
			260					265					270		
Gly	Ile	Ala	Met	Leu	Ile	Ile	Ser	Val	Leu	Ile	Ala	Ile	Tyr	Tyr	Asn
	275						280					285			
Val	Ile	Ile	Cys	Tyr	Thr	Leu	Phe	Tyr	Leu	Phe	Ala	Ser	Phe	Val	Ser
	290					295					300				
Val	Leu	Pro	Trp	Gly	Ser	Cys	Asn	Asn	Pro	Trp	Asn	Thr	Pro	Glu	Cys
	305				310					315				320	
Lys	Asp	Lys	Thr	Lys	Leu	Leu	Leu	Asp	Ser	Cys	Val	Ile	Ser	Asp	His
			325						330					335	
Pro	Lys	Ile	Gln	Ile	Lys	Asn	Ser	Thr	Phe	Cys	Met	Thr	Ala	Tyr	Pro
			340					345					350		
Asn	Val	Thr	Met	Val	Asn	Phe	Thr	Ser	Gln	Ala	Asn	Lys	Thr	Phe	Val
	355					360						365			
Ser	Gly	Ser	Glu	Glu	Tyr	Phe	Lys	Tyr	Phe	Val	Leu	Lys	Ile	Ser	Ala

- 56 -

370		375		380
Gly Ile Glu Tyr Pro	Gly Glu Ile Arg Trp Pro	Leu Ala Leu Cys Leu		
385	390	395	400	
Phe Leu Ala Trp Val	Ile Val Tyr Ala Ser Leu	Ala Lys Gly Ile Lys		
	405	410	415	
Thr Ser Gly Lys Val	Val Tyr Phe Thr Ala Thr	Phe Pro Tyr Val Val		
	420	425	430	
Leu Val Ile Leu Leu	Ile Arg Gly Val Thr Leu	Pro Gly Ala Gly Ala		
	435	440	445	
Gly Ile Trp Tyr Phe	Ile Thr Pro Lys Trp Glu	Lys Leu Thr Asp Ala		
	450	455	460	
Thr Val Trp Lys Asp	Ala Ala Thr Gln Ile Phe	Phe Ser Leu Ser Ala		
465	470	475	480	
Ala Trp Gly Gly Leu	Ile Thr Leu Ser Ser Tyr	Asn Lys Phe His Asn		
	485	490	495	
Asn Cys Tyr Arg Asp	Thr Leu Ile Val Thr Cys	Thr Asn Ser Ala Thr		
	500	505	510	
Ser Ile Phe Ala Gly	Phe Val Ile Phe Ser Val	Ile Gly Phe Met Ala		
	515	520	525	
Asn Glu Arg Lys Val	Asn Ile Glu Asn Val Ala	Asp Gln Gly Pro Gly		
	530	535	540	
Ile Ala Phe Val Val	Tyr Pro Glu Ala Leu Thr	Arg Leu Pro Leu Ser		
545	550	555	560	
Pro Phe Trp Ala Ile	Ile Phe Phe Leu Met Leu	Leu Thr Leu Gly Leu		
	565	570	575	
Asp Thr Met Phe Ala	Thr Ile Glu Thr Ile Val	Thr Ser Ile Ser Asp		
	580	585	590	
Glu Phe Pro Lys Tyr	Leu Arg Thr His Lys Pro	Val Phe Thr Leu Gly		
	595	600	605	
Cys Cys Ile Cys Phe	Phe Ile Met Gly Phe Pro	Met Ile Thr Gln Gly		
	610	615	620	
Gly Ile Tyr Met Phe	Gln Leu Val Asp Thr Tyr	Ala Ala Ser Tyr Ala		
625	630	635	640	
Leu Val Ile Ile Ala	Ile Phe Glu Leu Val Gly	Ile Ser Tyr Val Tyr		
	645	650	655	
Gly Leu Gln Arg Phe	Cys Glu Asp Ile Glu	Met Met Ile Gly Phe Gln		
	660	665	670	
Pro Asn Ile Phe Trp	Lys Val Cys Trp Ala Phe	Val Thr Pro Thr Ile		
	675	680	685	
Leu Thr Phe Ile Leu	Cys Phe Ser Phe Tyr Gln	Trp Glu Pro Met Thr		
	690	695	700	
Tyr Gly Ser Tyr Arg	Tyr Pro Asn Trp Ser Met	Val Leu Gly Trp Leu		
705	710	715	720	
Met Leu Ala Cys Ser	Val Ile Trp Ile Pro Ile	Met Phe Val Ile Lys		
	725	730	735	
Met His Leu Ala Pro	Gly Arg Phe Ile Glu	Arg Leu Lys Leu Val Cys		
	740	745	750	
Ser Pro Gln Pro Asp	Trp Gly Pro Phe Leu	Ala Gln His Arg Gly Glu		
	755	760	765	
Arg Tyr Lys Asn Met	Ile Asp Pro Leu Gly	Thr Ser Ser Leu Gly Leu		
	770	775	780	
Lys Leu Pro Val Lys	Asp Leu Glu Leu Gly	Thr Gln Cys		
785	790	795		

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 2397 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: Coding Sequence

(B) LOCATION: 1...2391

(D) OTHER INFORMATION:

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

- 57 -

ATG GAT TGC AGT GCT CCC AAG GAA ATG AAT AAA CTG CCA GCC AAC AGC Met Asp Cys Ser Ala Pro Lys Glu Met Asn Lys Leu Pro Ala Asn Ser 1 5 10 15	48
CCG GAG GCG GCG GCG GCG CAG GGC CAC CCG GAT GGC CCA TGC GCT CCC Pro Glu Ala Ala Ala Ala Gln Gly His Pro Asp Gly Pro Cys Ala Pro 20 25 30	96
AGG ACG AGC CCG GAG CAG GAG CTT CCC GCG GCT GCC GCC CCG CCG CCG Arg Thr Ser Pro Glu Gln Glu Leu Pro Ala Ala Ala Ala Pro Pro Pro 35 40 45	144
CCA CGT GTG CCC AGG TCC GCT TCC ACC GGC GCC CAA ACT TTC CAG TCA Pro Arg Val Pro Arg Ser Ala Ser Thr Gly Ala Gln Thr Phe Gln Ser 50 55 60	192
GCG GAC GCG CGA GCC TGC GAG GCT GAG CGG CCA GGA GTG GGG TCT TGC Ala Asp Ala Arg Ala Cys Glu Ala Glu Arg Pro Gly Val Gly Ser Cys 65 70 75 80	240
AAA CTC AGT AGC CCG CGG GCG CAG GCG GCC TCT GCA GCT CTG CGG GAC Lys Leu Ser Ser Pro Arg Ala Gln Ala Ala Ser Ala Ala Leu Arg Asp 85 90 95	288
TTG AGA GAG GCG CAA GGC GCG CAG GCC TCG CCC CCT CCC GGG AGC TCC Leu Arg Glu Ala Gln Gly Ala Gln Ala Ser Pro Pro Pro Gly Ser Ser 100 105 110	336
GGG CCC GGC AAC GCG CTG CAC TGT AAG ATC CCT TCT CTG CGA GGC CCG Gly Pro Gly Asn Ala Leu His Cys Lys Ile Pro Ser Leu Arg Gly Pro 115 120 125	384
GAG GGG GAT GCG AAC GTG AGT GTG GGC AAG GGC ACC CTG GAG CGG AAC Glu Gly Asp Ala Asn Val Ser Val Gly Lys Gly Thr Leu Glu Arg Asn 130 135 140	432
AAT ACC CCT GTT GTG GGC TGG GTG AAC ATG AGC CAG AGC ACC GTG GTG Asn Thr Pro Val Val Gly Trp Val Asn Met Ser Gln Ser Thr Val Val 145 150 155 160	480
CTG GGC ACG GAT GGA ATC ACG TCC GTG CTC CCG GGC AGC GTG GCC ACC Leu Gly Thr Asp Gly Ile Thr Ser Val Leu Pro Gly Ser Val Ala Thr 165 170 175	528
GTT GCC ACC CAG GAG GAC GAG CAA GGG GAT GAG AAT AAG GCC CGA GGG Val Ala Thr Gln Glu Asp Glu Gln Gly Asp Glu Asn Lys Ala Arg Gly 180 185 190	576
AAC TGG TCC AGC AAA CTG GAC TTC ATC CTG TCC ATG GTG GGG TAC GCA Asn Trp Ser Ser Lys Leu Asp Phe Ile Leu Ser Met Val Gly Tyr Ala 195 200 205	624
GTG GGG CTG GGC AAT GTC TGG AGG TTT CCC TAC CTG GCC TTC CAG AAC Val Gly Leu Gly Asn Val Trp Arg Phe Pro Tyr Leu Ala Phe Gln Asn 210 215 220	672
GGG GGA GGT GCT TTC CTC ATC CCT TAC CTG ATG ATG CTG GCT CTG GCT Gly Gly Gly Ala Phe Leu Ile Pro Tyr Leu Met Met Leu Ala Leu Ala 225 230 235 240	720
GGA TTA CCC ATC TTC TTC TTG GAG GTG TCG CTG GGC CAG TTT GCC AGC Gly Leu Pro Ile Phe Phe Leu Glu Val Ser Leu Gly Gln Phe Ala Ser 245 250 255	768
CAG GGA CCA GTG TCT GTG TGG AAG GCC ATC CCA GCT CTA CAA GGC TGT	816

- 58 -

Gln	Gly	Pro	Val	Ser	Val	Trp	Lys	Ala	Ile	Pro	Ala	Leu	Gln	Gly	Cys	
			260					265					270			
GGC	ATC	GCG	ATG	CTG	ATC	ATC	TCT	GTC	CTA	ATA	GCC	ATA	TAC	TAC	AAT	864
Gly	Ile	Ala	Met	Leu	Ile	Ile	Ser	Val	Leu	Ile	Ala	Ile	Tyr	Tyr	Asn	
		275					280					285				
GTG	ATT	ATT	TGC	TAT	ACA	CTT	TTC	TAC	CTG	TTT	GCC	TCC	TTT	GTG	TCT	912
Val	Ile	Ile	Cys	Tyr	Thr	Leu	Phe	Tyr	Leu	Phe	Ala	Ser	Phe	Val	Ser	
	290					295					300					
GTA	CTA	CCC	TGG	GGC	TCC	TGC	AAC	AAC	CCT	TGG	AAT	ACG	CCA	GAA	TGC	960
Val	Leu	Pro	Trp	Gly	Ser	Cys	Asn	Asn	Pro	Trp	Asn	Thr	Pro	Glu	Cys	
305					310					315					320	
AAA	GAT	AAA	ACC	AAA	CTT	TTA	TTA	GAT	TCC	TGT	GTT	ATC	AGT	GAC	CAT	1008
Lys	Asp	Lys	Thr	Lys	Leu	Leu	Leu	Asp	Ser	Cys	Val	Ile	Ser	Asp	His	
				325					330					335		
CCC	AAA	ATA	CAG	ATC	AAG	AAC	TCG	ACT	TTC	TGC	ATG	ACC	GCT	TAT	CCC	1056
Pro	Lys	Ile	Gln	Ile	Lys	Asn	Ser	Thr	Phe	Cys	Met	Thr	Ala	Tyr	Pro	
			340					345					350			
AAC	GTG	ACA	ATG	GTT	AAT	TTC	ACC	AGC	CAG	GCC	AAT	AAG	ACA	TTT	GTC	1104
Asn	Val	Thr	Met	Val	Asn	Phe	Thr	Ser	Gln	Ala	Asn	Lys	Thr	Phe	Val	
		355				360						365				
AGT	GGA	AGT	GAG	GAG	TAC	TTC	AAG	TAC	TTT	GTG	CTG	AAG	ATT	TCT	GCA	1152
Ser	Gly	Ser	Glu	Glu	Tyr	Phe	Lys	Tyr	Phe	Val	Leu	Lys	Ile	Ser	Ala	
	370					375					380					
GGG	ATT	GAA	TAT	CCT	GGC	GAG	ATC	AGG	TGG	CCA	CTA	GCT	CTC	TGC	CTC	1200
Gly	Ile	Glu	Tyr	Pro	Gly	Glu	Ile	Arg	Trp	Pro	Leu	Ala	Leu	Cys	Leu	
385					390					395					400	
TTC	CTG	GCT	TGG	GTC	ATT	GTG	TAT	GCA	TCG	TTG	GCT	AAA	GGA	ATC	AAG	1248
Phe	Leu	Ala	Trp	Val	Ile	Val	Tyr	Ala	Ser	Leu	Ala	Lys	Gly	Ile	Lys	
				405					410					415		
ACT	TCA	GGA	AAA	GTG	GTG	TAC	TTC	ACG	GCC	ACG	TTC	CCG	TAT	GTC	GTA	1296
Thr	Ser	Gly	Lys	Val	Val	Tyr	Phe	Thr	Ala	Thr	Phe	Pro	Tyr	Val	Val	
			420					425					430			
CTC	GTG	ATC	CTC	CTC	ATC	CGA	GGA	GTC	ACC	CTG	CCT	GGA	GCT	GGA	GCT	1344
Leu	Val	Ile	Leu	Leu	Ile	Arg	Gly	Val	Thr	Leu	Pro	Gly	Ala	Gly	Ala	
		435					440					445				
GGG	ATC	TGG	TAC	TTC	ATC	ACA	CCC	AAG	TGG	GAG	AAA	CTC	ACG	GAT	GCC	1392
Gly	Ile	Trp	Tyr	Phe	Ile	Thr	Pro	Lys	Trp	Glu	Lys	Leu	Thr	Asp	Ala	
	450					455					460					
ACG	GTG	TGG	AAA	GAT	GCT	GCC	ACT	CAG	ATT	TTC	TTC	TCT	TTA	TCT	GCT	1440
Thr	Val	Trp	Lys	Asp	Ala	Ala	Thr	Gln	Ile	Phe	Phe	Ser	Leu	Ser	Ala	
	465				470				475					480		
GCA	TGG	GGA	GGC	CTG	ATC	ACT	CTC	TCT	TCT	TAC	AAC	AAA	TTC	CAC	AAC	1488
Ala	Trp	Gly	Gly	Leu	Ile	Thr	Leu	Ser	Ser	Tyr	Asn	Lys	Phe	His	Asn	
				485				490						495		
AAC	TGC	TAC	AGG	GAC	ACT	CTA	ATT	GTC	ACC	TGC	ACC	AAC	AGT	GCC	ACA	1536
Asn	Cys	Tyr	Arg	Asp	Thr	Leu	Ile	Val	Thr	Cys	Thr	Asn	Ser	Ala	Thr	
			500					505					510			
AGC	ATC	TTT	GCC	GGC	TTC	GTC	ATC	TTC	TCC	GTT	ATC	GGC	TTC	ATG	GCC	1584
Ser	Ile	Phe	Ala	Gly	Phe	Val	Ile	Phe	Ser	Val	Ile	Gly	Phe	Met	Ala	

- 59 -

515	520	525	
AAT GAA CGC AAA GTC AAC ATT GAG AAT GTG GCA GAC CAA GGG CCA GGC Asn Glu Arg Lys Val Asn Ile Glu Asn Val Ala Asp Gln Gly Pro Gly 530 535 540			1632
ATT GCA TTT GTG GTT TAC CCG GAA GCC TTA ACC AGG CTG CCT CTC TCT Ile Ala Phe Val Val Tyr Pro Glu Ala Leu Thr Arg Leu Pro Leu Ser 545 550 555 560			1680
CCG TTC TGG GCC ATC ATC TTT TTC CTG ATG CTC CTC ACT CTT GGA CTT Pro Phe Trp Ala Ile Ile Phe Phe Leu Met Leu Leu Thr Leu Gly Leu 565 570 575			1728
GAC ACT ATG TTT GCC ACC ATC GAG ACC ATA GTG ACC TCC ATC TCA GAC Asp Thr Met Phe Ala Thr Ile Glu Thr Ile Val Thr Ser Ile Ser Asp 580 585 590			1776
GAG TTT CCC AAG TAC CTA CGC ACA CAC AAG CCA GTG TTT ACT CTG GGC Glu Phe Pro Lys Tyr Leu Arg Thr His Lys Pro Val Phe Thr Leu Gly 595 600 605			1824
TGC TGC ATT TGT TTC TTC ATC ATG GGT TTT CCA ATG ATC ACT CAG GGT Cys Cys Ile Cys Phe Phe Ile Met Gly Phe Pro Met Ile Thr Gln Gly 610 615 620			1872
GGA ATT TAC ATG TTT CAG CTT GTG GAC ACC TAT GCT GCC TCC TAT GCC Gly Ile Tyr Met Phe Gln Leu Val Asp Thr Tyr Ala Ala Ser Tyr Ala 625 630 635 640			1920
CTT GTC ATC ATT GCC ATT TTT GAG CTC GTG GGG ATC TCT TAT GTG TAT Leu Val Ile Ile Ala Ile Phe Glu Leu Val Gly Ile Ser Tyr Val Tyr 645 650 655			1968
GGC TTG CAA AGA TTC TGT GAA GAT ATA GAG ATG ATG ATT GGA TTC CAG Gly Leu Gln Arg Phe Cys Glu Asp Ile Glu Met Met Ile Gly Phe Gln 660 665 670			2016
CCT AAC ATC TTC TGG AAA GTC TGC TGG GCA TTT GTA ACC CCA ACC ATT Pro Asn Ile Phe Trp Lys Val Cys Trp Ala Phe Val Thr Pro Thr Ile 675 680 685			2064
TTA ACC TTT ATC CTT TGC TTC AGC TTT TAC CAG TGG GAG CCC ATG ACC Leu Thr Phe Ile Leu Cys Phe Ser Phe Tyr Gln Trp Glu Pro Met Thr 690 695 700			2112
TAT GGC TCT TAC CGC TAT CCT AAC TGG TCC ATG GTG CTC GGA TGG CTA Tyr Gly Ser Tyr Arg Tyr Pro Asn Trp Ser Met Val Leu Gly Trp Leu 705 710 715 720			2160
ATG CTC GCC TGT TCC GTC ATC TGG ATC CCA ATT ATG TTT GTG ATA AAA Met Leu Ala Cys Ser Val Ile Trp Ile Pro Ile Met Phe Val Ile Lys 725 730 735			2208
ATG CAT CTG GCC CCT GGA AGA TTT ATT GAG AGG CTG AAG TTG GTG TGC Met His Leu Ala Pro Gly Arg Phe Ile Glu Arg Leu Lys Leu Val Cys 740 745 750			2256
TCG CCA CAG CCG GAC TGG GGC CCA TTC TTA GCT CAA CAC CGC GGG GAG Ser Pro Gln Pro Asp Trp Gly Pro Phe Leu Ala Gln His Arg Gly Glu 755 760 765			2304
CGT TAC AAG AAC ATG ATC GAC CCC TTG GGA ACC TCT TCC TTG GGA CTC Arg Tyr Lys Asn Met Ile Asp Pro Leu Gly Thr Ser Ser Leu Gly Leu 770 775 780			2352

- 60 -

AAA CTG CCA GTG AAG GAT TTG GAA CTG GGA ACG CAA TGC TAATCC
 Lys Leu Pro Val Lys Asp Leu Glu Leu Gly Thr Gln Cys
 785 790 795

2397

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 949 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

AGGCGCAAGG	CGCGCAGGCC	TCGCCCCCTC	CCGGGAGCTC	CGGGCCCGGC	AACGCGTTGC	60
ACTGTAAGAT	CCCTTCTCTG	CGAGGCCCCG	AGGGGGATGC	GAACGTGAGT	GTGGGCAAGG	120
GCACCCTGGA	GCGGAACAAT	ACCCCTGTTG	TGGGTGGGT	GAACATGAGC	CAGAGCACCG	180
TGGTGCTGGG	CACGGATGGA	ATCACGTCCG	TGCTCCCGGG	CAGCGTGGCC	ACCGTTGCCA	240
CCGAGGAGGA	CGAGCAAGGG	GATGAGAATA	AGGCCCGAGG	GAAGTGGTCC	AGCAAAGTGG	300
ACTTCATCCT	GTCCATGGTG	GGGTACGCAG	TGGGGCTGGG	CAATGTCTGG	AGGTTCCCT	360
ACCTGGCCTT	CCAGAACGGG	GGAGGTGCTT	TCCTCATCCC	TTACCTGATG	ATGCTGGCTC	420
TGGCTGGATT	ACCCATCTTC	TTCTTGGAGG	TGTCGCTGGG	CCAGTTTGCC	AGCCAGGGAC	480
CAGTGTCTGT	GTGGAAGGCC	ATCCCAGCTC	TACAAGGCTG	TGGCATCGCG	ATGCTGATCA	540
TCTCTGTCCT	AATAGCCATA	TACTACAATG	TGATTATTTG	CTATACACTT	TTCTACCTGT	600
TTGCCTCCTT	TGTGTCTGTA	CTACCCTGGG	GCTCCTGCAA	CAACCCTTGG	AATACACCAG	660
AATGCAAAGA	TAAACCAAA	CTTTTATTAG	ATTCTGTGT	TATCAGTGAC	CATCCCCAAA	720
TACAGATCAA	GAAGTCGACT	TTCTGCATGA	CCGCTTATCC	CAACGTGACA	ATGGTTAATT	780
TCACCAGCCA	GGCCAATAAG	ACATTTGTCA	GTGGAAGTGA	AGAGTACTTC	AAGTACTTTG	840
TGCTGAAGAT	TTCTGCAGGG	ATTGAATATC	CTGGCGAGAT	CAGGTGGCCA	CTAGCTCTCT	900
GCCTCTTCCT	GGCTTGGGTC	ATTGTGTATG	CATCGTTGGC	TAAAGGAAT		949

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 315 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

Ala	Gln	Gly	Ala	Gln	Ala	Ser	Pro	Pro	Pro	Gly	Ser	Ser	Gly	Pro	Gly
1				5					10					15	
Asn	Ala	Leu	His	Cys	Lys	Ile	Pro	Ser	Leu	Arg	Gly	Pro	Glu	Gly	Asp
			20					25					30		
Ala	Asn	Val	Ser	Val	Gly	Lys	Gly	Thr	Leu	Glu	Arg	Asn	Asn	Thr	Pro
		35				40						45			
Val	Val	Gly	Trp	Val	Asn	Met	Ser	Gln	Ser	Thr	Val	Val	Leu	Gly	Thr
	50				55						60				
Asp	Gly	Ile	Thr	Ser	Val	Leu	Pro	Gly	Ser	Val	Ala	Thr	Val	Ala	Thr
65					70				75					80	
Gln	Glu	Asp	Glu	Gln	Gly	Asp	Glu	Asn	Lys	Ala	Arg	Gly	Asn	Trp	Ser
			85					90					95		
Ser	Lys	Leu	Asp	Phe	Ile	Leu	Ser	Met	Val	Gly	Tyr	Ala	Val	Gly	Leu
			100					105					110		
Gly	Asn	Val	Trp	Arg	Phe	Pro	Tyr	Leu	Ala	Phe	Gln	Asn	Gly	Gly	Gly
		115					120					125			
Ala	Phe	Leu	Ile	Pro	Tyr	Leu	Met	Met	Leu	Ala	Leu	Ala	Gly	Leu	Pro
	130					135					140				
Ile	Phe	Phe	Leu	Glu	Val	Ser	Leu	Gly	Gln	Phe	Ala	Ser	Gln	Gly	Pro
145					150				155					160	
Val	Ser	Val	Trp	Lys	Ala	Ile	Pro	Ala	Leu	Gln	Gly	Cys	Gly	Ile	Ala
			165					170					175		
Met	Leu	Ile	Ile	Ser	Val	Leu	Ile	Ala	Ile	Tyr	Tyr	Asn	Val	Ile	Ile
			180					185					190		
Cys	Tyr	Thr	Leu	Phe	Tyr	Leu	Phe	Ala	Ser	Phe	Val	Ser	Val	Leu	Pro
		195					200					205			
Trp	Gly	Ser	Cys	Asn	Asn	Pro	Trp	Asn	Thr	Pro	Glu	Cys	Lys	Asp	Lys

- 61 -

210	215	220
Thr Lys Leu Leu Leu Asp	Ser Cys Val Ile Ser	Asp His Pro Lys Ile
225	230	235
Gln Ile Lys Asn Ser Thr	Phe Cys Met Thr Ala	Tyr Pro Asn Val Thr
245	250	255
Met Val Asn Phe Thr Ser	Gln Ala Asn Lys Thr	Phe Val Ser Gly Ser
260	265	270
Glu Glu Tyr Phe Lys Tyr	Phe Val Leu Lys Ile	Ser Ala Gly Ile Glu
275	280	285
Tyr Pro Gly Glu Ile Arg	Trp Pro Leu Ala Leu	Cys Leu Phe Leu Ala
290	295	300
Trp Val Ile Val Tyr Ala	Ser Leu Ala Lys Gly	
305	310	315

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 949 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

AGGCGCAAGG	CGCGCAGGCC	TCGCCCCCTC	CCGGGAGCTC	CGGGCCCGGC	AACGCGTTGC	60
ACTGTAAGAT	CCCTTCTCTG	CGAGGCCCGG	AGGGGGATGC	GAACGTGAGT	GTGGGCAAGG	120
GCACCCTGGA	GCGGAACAAT	ACCCCTGTTG	TGGGCTGGGT	GAACATGAGC	CAGAGCACCG	180
TGGTGCTGGG	CACGGATGGA	ATCACGTCCG	TGCTCCCGGG	CAGCGTGGCC	ACCGTTGCCA	240
CCCAGGAGGA	CGAGCAAGGG	GATGAGAATA	AGGCCGAGG	GAACGGTCC	AGCAAACCTGG	300
ACTTCATCCT	GTCCATGGTG	GGGTACGCAG	TGGGGCTGGG	CAATGTCTGG	AGGTTTCCCT	360
ACCTGGCCTT	CCAGAACGGG	GGAGGTGCTT	TCCTCATCCC	TTACCTGATG	ATGCTGGCTC	420
TGGCTGGATT	ACCCATCCTC	TTCTTGGAGG	TGTCGCTGGG	CCAGTTTGCC	AGCCAGGGAC	480
CAGTGTCTGT	GTGGAAGGCC	ATCCCAGCTC	TACAAGGCTG	TGGCATCGCG	ATGCTGATCA	540
TCTCTGTCCT	AATAGCCATA	TACTACAATG	TGATTATTG	CTATACACTT	TTCTACCTGT	600
TTGCCTCCTT	TGTGTCTGTA	CTACCCTGGG	GCTCCTGCAA	CAACCCTTGG	AATACACCAG	660
AATGCAAAGA	TAAAACCAAA	CTTTTATTAG	ATTCTGTGT	TATCAGTGAC	CATCCCCAAA	720
TACAGATCAA	GAACGCGACT	TTCTGCATGA	CCGCTTATCC	CAACGTGACA	ATGGTTAATT	780
TCACCAGCCA	GGCCAATAAG	ATATTTGTCA	GTGGAAGTGA	AGAGTACTTC	AAGTACTTTG	840
TGCTGAAGAT	TTCTGCAGGG	ATTGAATATC	CTGGCGAGAT	CAGGTGGCCA	CTAGCTCTCT	900
GCCTCTTCCT	GGCTTGGGTC	ATTGTGTATG	CATCGTTGGC	TAAAGGAAT		949

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 315 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

Ala	Gln	Gly	Ala	Gln	Ala	Ser	Pro	Pro	Pro	Gly	Ser	Ser	Gly	Pro	Gly
1				5					10					15	
Asn	Ala	Leu	His	Cys	Lys	Ile	Pro	Ser	Leu	Arg	Gly	Pro	Glu	Gly	Asp
			20					25					30		
Ala	Asn	Val	Ser	Val	Gly	Lys	Gly	Thr	Leu	Glu	Arg	Asn	Asn	Thr	Pro
		35					40					45			
Val	Val	Gly	Trp	Val	Asn	Met	Ser	Gln	Ser	Thr	Val	Val	Leu	Gly	Thr
	50				55						60				
Asp	Gly	Ile	Thr	Ser	Val	Leu	Pro	Gly	Ser	Val	Ala	Thr	Val	Ala	Thr
65					70				75					80	
Gln	Glu	Asp	Glu	Gln	Gly	Asp	Glu	Asn	Lys	Ala	Arg	Gly	Asn	Trp	Ser
			85						90					95	
Ser	Lys	Leu	Asp	Phe	Ile	Leu	Ser	Met	Val	Gly	Tyr	Ala	Val	Gly	Leu
		100						105					110		
Gly	Asn	Val	Trp	Arg	Phe	Pro	Tyr	Leu	Ala	Phe	Gln	Asn	Gly	Gly	Gly
	115					120					125				
Ala	Phe	Leu	Ile	Pro	Tyr	Leu	Met	Met	Leu	Ala	Leu	Ala	Gly	Leu	Pro
	130					135					140				

- 62 -

Ile Leu Phe Leu Glu Val Ser Leu Gly Gln Phe Ala Ser Gln Gly Pro
 145 150 155 160
 Val Ser Val Trp Lys Ala Ile Pro Ala Leu Gln Gly Cys Gly Ile Ala
 165 170 175
 Met Leu Ile Ile Ser Val Leu Ile Ala Ile Tyr Tyr Asn Val Ile Ile
 180 185 190
 Cys Tyr Thr Leu Phe Tyr Leu Phe Ala Ser Phe Val Ser Val Leu Pro
 195 200 205
 Trp Gly Ser Cys Asn Asn Pro Trp Asn Thr Pro Glu Cys Lys Asp Lys
 210 215 220
 Thr Lys Leu Leu Leu Asp Ser Cys Val Ile Ser Asp His Pro Lys Ile
 225 230 235 240
 Gln Ile Lys Asn Ser Thr Phe Cys Met Thr Ala Tyr Pro Asn Val Thr
 245 250 255
 Met Val Asn Phe Thr Ser Gln Ala Asn Lys Ile Phe Val Ser Gly Ser
 260 265 270
 Glu Glu Tyr Phe Lys Tyr Phe Val Leu Lys Ile Ser Ala Gly Ile Glu
 275 280 285
 Tyr Pro Gly Glu Ile Arg Trp Pro Leu Ala Leu Cys Leu Phe Leu Ala
 290 295 300
 Trp Val Ile Val Tyr Ala Ser Leu Ala Lys Gly
 305 310 315

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 949 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: double

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

```

AGGCGCAAGG CGCGCAGGCC TCGCCCCCTC CCGGGAGCTC CGGGCCGGGC AACGCGCTGC      60
ACTGTAAGAT CCCTTCTCTG CGAGGCCCGG AGGGGGATGC GAACGTGAGT GTGGGCAAGG      120
GCACCTTGGA GCGGAACAAT ACCCTGTGTG TGGGCTGGGT GAACATGAGC CAGAGCACCG      180
TGGTGCTGGG CACGGATGGA ATCACGTCCG TGCTCCCGGG CAGCGTGGCC ACCGTTGCCA      240
CCCAGGAGGA CGAGCAAGGG GATGAGAATA AGGCCCGAGG GAACTGGTCC AGCAAAGTGG      300
ACTTCATCCT GTCCATGGTG GGGTACGCAG TGGGGCTGGG CAATGTCTGG AGGTTTCCTT      360
ACCTGGCCTT CCAGAACGGG GGAGGTGCTT TCCTCATCCC TTACCTGATG ATGCTGGCTC      420
TGGCTGGATT ACCCATCTTC TTCTTGGAGG TGTCGCTGGG CCAGTTTGCC AGCCAGGGAC      480
CGGTGTCTGT GTGGAAGGCC ATCCCAGCTC TACAAGGCTG TGGCATCGCG ATGCTGATCA      540
TCTCTGTCCT AATAGCCATA TACTACAATG TGATTATTTG CTATACACTT TTCTACCTGT      600
TTGCCTCCTT TGTGTCTGTA CTACCCCTGG GCTCCTGCAA CAACCCTTGG AATACGCCAG      660
AATGCAAAGA TAAACCAAAA CTTTATTAG ATTCTGTGT TATCAGTGAC CATCCCAAAA      720
TACAGATCAA GAACTCGACT TTCTGCATGA CCGCTTATCC CAACGTGACA ATGGTTAATT      780
TCACCAGCCA GGCCAATAAG ACATTTGTCA GTGGAAGTGA AGAGTACTTC AAGTACTTTG      840
TGCTGAAGAT TTCTGCAGGG ATTGAATATC CTGGCGAGAT CAGGTGGCCA CTAGCTCTCT      900
GCCCCCTCCT GGCTTGGGTC ATTGTGTATG CATCGTTGGC TAAAGGAAT      949

```

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 315 amino acids

(B) TYPE: amino acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

Ala Gln Gly Ala Gln Ala Ser Pro Pro Pro Gly Ser Ser Gly Pro Gly
 1 5 10 15
 Asn Ala Leu His Cys Lys Ile Pro Ser Leu Arg Gly Pro Glu Gly Asp
 20 25 30
 Ala Asn Val Ser Val Gly Lys Gly Thr Leu Glu Arg Asn Asn Thr Pro
 35 40 45
 Val Val Gly Trp Val Asn Met Ser Gln Ser Thr Val Val Leu Gly Thr
 50 55 60
 Asp Gly Ile Thr Ser Val Leu Pro Gly Ser Val Ala Thr Val Ala Thr

- 63 -

65					70					75				80
Gln	Glu	Asp	Glu	Gln	Gly	Asp	Glu	Asn	Lys	Ala	Arg	Gly	Asn	Trp
				85					90					95
Ser	Lys	Leu	Asp	Phe	Ile	Leu	Ser	Met	Val	Gly	Tyr	Ala	Val	Gly
			100					105					110	
Gly	Asn	Val	Trp	Arg	Phe	Pro	Tyr	Leu	Ala	Phe	Gln	Asn	Gly	Gly
		115					120					125		
Ala	Phe	Leu	Ile	Pro	Tyr	Leu	Met	Met	Leu	Ala	Leu	Ala	Gly	Leu
		130				135					140			
Ile	Phe	Phe	Leu	Glu	Val	Ser	Leu	Gly	Gln	Phe	Ala	Ser	Gln	Gly
145					150					155				160
Val	Ser	Val	Trp	Lys	Ala	Ile	Pro	Ala	Leu	Gln	Gly	Cys	Gly	Ile
				165					170					175
Met	Leu	Ile	Ile	Ser	Val	Leu	Ile	Ala	Ile	Tyr	Tyr	Asn	Val	Ile
			180					185					190	
Cys	Tyr	Thr	Leu	Phe	Tyr	Leu	Phe	Ala	Ser	Phe	Val	Ser	Val	Leu
		195					200					205		
Trp	Gly	Ser	Cys	Asn	Asn	Pro	Trp	Asn	Thr	Pro	Glu	Cys	Lys	Asp
		210				215					220			
Thr	Lys	Leu	Leu	Leu	Asp	Ser	Cys	Val	Ile	Ser	Asp	His	Pro	Lys
225					230					235				240
Gln	Ile	Lys	Asn	Ser	Thr	Phe	Cys	Met	Thr	Ala	Tyr	Pro	Asn	Val
			245						250					255
Met	Val	Asn	Phe	Thr	Ser	Gln	Ala	Asn	Lys	Thr	Phe	Val	Ser	Gly
			260					265					270	
Glu	Glu	Tyr	Phe	Lys	Tyr	Phe	Val	Leu	Lys	Ile	Ser	Ala	Gly	Ile
		275					280					285		
Tyr	Pro	Gly	Glu	Ile	Arg	Trp	Pro	Leu	Ala	Leu	Cys	Pro	Phe	Leu
		290				295					300			
Trp	Val	Ile	Val	Tyr	Ala	Ser	Leu	Ala	Lys	Gly				
305					310					315				

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1303 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

AGGCGCAAAG	CGCGCAGGCC	TCGCCCCCTC	CCGGGAGCTC	CGGGCCCGGC	AACGCGCTGC	60
ACTGTAAGAT	CCCTTCTCTG	CGAGGCCCGG	AGGGGGATGC	GAACGTGAGT	GTGGGCAAGG	120
GCACCTGGA	GCGGAACAAT	ACCCCTGTTG	TGGGCTGGGT	GAACATGAGC	CAGAGCACCG	180
TGGTGTGGG	CACGGATGGA	ATCACGTCCG	TGCTCCCGGG	CAGCGTGGCC	ACCGTTGCCA	240
CCCAGGAGGA	CGAGCAAGGG	GATGAGAATA	AGGCCCCGAGG	GAAGTGGTCC	AGCAAAGTGG	300
ACTTCATCCT	GTCCATGGTG	GGGTACGCAG	TGGGGCTGGG	CAATGTCTGG	AGGTTTCCCT	360
ACCTGGCCTT	CCAGAACGGG	GGAGGTGCTT	TCCTCATCCC	TTACCTGATG	ATGCTGGCTC	420
TGGCTGGATT	ACCCATCTTC	TTCTTGGAGG	TGTCGCTGGG	CCAGTTTGCC	AGCCAGGGAC	480
CAGTGCTGT	GTGGAAGGCC	ATCCCAGCTC	TACAAGGCTG	TGGCATCGCG	ATGCTGATCA	540
TCTCTGTCCT	AATAGCCATA	TACTACAATG	TGATTATTTG	CTATACACTT	TTCTACCTGT	600
TTGCCTCCTT	TGTGTCTCTA	CTACCCTGGG	GCTCCTGCAA	CAACCCTTGG	AATACGCCAG	660
AATGCAAAGA	TAAAACCAAA	CTTTTATTAG	ATTCTGTGT	TATCAGTGAC	CATCCCAAAA	720
TACAGATCAA	GAAGTCGACT	TTCTGCATGA	CCGCTTATCC	CAACGTGACA	ATGGTTAATT	780
TCACCAGCCA	GGCCAATAAG	ACATTTGTCA	GTGGAAGTGA	AGAGTACTTC	AAGTACTTTG	840
TGCTGAAGAT	TTCTGCAGGG	ATTGAATATC	CTGGCGAGAT	CAGGTGGCCA	CTAGCTCTCT	900
GCCTCTCCT	GGCTTGGGTC	ATTGTGTATG	CATCGTTGGC	TAAAGGAATC	AAGACTTCAG	960
GAAAAGTGGT	GTAATTCACG	GCCACGTTCC	CGTATGTCGT	ACTCGTGATC	CTCCTCATCC	1020
GAGGAGTCAC	CCTGCCTGGA	GCTGGAGCTG	GGATCTGGTA	CTTCATCACA	CCCAAGTGGG	1080
AGAAATCCAC	GGATGCCACG	GTGTGGAAG	ATGCTGCCAC	TCAGATTTTC	TTCTCTTTAT	1140
CTGCTGCATG	GGGAGGCCCTG	ATCACTCTCT	CTTCTTACAA	CAAATCCAC	AACAAGTGGT	1200
ACAGGGACAT	TCTAATTGTC	ACCTGCACCA	ACAGTGCCAC	AAGCATCTTT	GCCGGCTTCG	1260
TCATCTTCTC	CGTTATCGGC	TTCATGGCCA	ATGAACGCAA	AGT		1303

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:

- 64 -

- (A) LENGTH: 433 amino acids
 (B) TYPE: amino acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

```

Ala Gln Ser Ala Gln Ala Ser Pro Pro Pro Gly Ser Ser Gly Pro Gly
 1      5      10      15
Asn Ala Leu His Cys Lys Ile Pro Ser Leu Arg Gly Pro Glu Gly Asp
 20      25      30
Ala Asn Val Ser Val Gly Lys Gly Thr Leu Glu Arg Asn Asn Thr Pro
 35      40      45
Val Val Gly Trp Val Asn Met Ser Gln Ser Thr Val Val Leu Gly Thr
 50      55      60
Asp Gly Ile Thr Ser Val Leu Pro Gly Ser Val Ala Thr Val Ala Thr
 65      70      75      80
Gln Glu Asp Glu Gln Gly Asp Glu Asn Lys Ala Arg Gly Asn Trp Ser
 85      90      95
Ser Lys Leu Asp Phe Ile Leu Ser Met Val Gly Tyr Ala Val Gly Leu
 100      105      110
Gly Asn Val Trp Arg Phe Pro Tyr Leu Ala Phe Gln Asn Gly Gly Gly
 115      120      125
Ala Phe Leu Ile Pro Tyr Leu Met Met Leu Ala Leu Ala Gly Leu Pro
 130      135      140
Ile Phe Phe Leu Glu Val Ser Leu Gly Gln Phe Ala Ser Gln Gly Pro
 145      150      155      160
Val Ser Val Trp Lys Ala Ile Pro Ala Leu Gln Gly Cys Gly Ile Ala
 165      170      175
Met Leu Ile Ile Ser Val Leu Ile Ala Ile Tyr Tyr Asn Val Ile Ile
 180      185      190
Cys Tyr Thr Leu Phe Tyr Leu Phe Ala Ser Phe Val Ser Leu Leu Pro
 195      200      205
Trp Gly Ser Cys Asn Asn Pro Trp Asn Thr Pro Glu Cys Lys Asp Lys
 210      215      220
Thr Lys Leu Leu Leu Asp Ser Cys Val Ile Ser Asp His Pro Lys Ile
 225      230      235      240
Gln Ile Lys Asn Ser Thr Phe Cys Met Thr Ala Tyr Pro Asn Val Thr
 245      250      255
Met Val Asn Phe Thr Ser Gln Ala Asn Lys Thr Phe Val Ser Gly Ser
 260      265      270
Glu Glu Tyr Phe Lys Tyr Phe Val Leu Lys Ile Ser Ala Gly Ile Glu
 275      280      285
Tyr Pro Gly Glu Ile Arg Trp Pro Leu Ala Leu Cys Leu Phe Leu Ala
 290      295      300
Trp Val Ile Val Tyr Ala Ser Leu Ala Lys Gly Ile Lys Thr Ser Gly
 305      310      315      320
Lys Val Val Tyr Phe Thr Ala Thr Phe Pro Tyr Val Val Leu Val Ile
 325      330      335
Leu Leu Ile Arg Gly Val Thr Leu Pro Gly Ala Gly Ala Gly Ile Trp
 340      345      350
Tyr Phe Ile Thr Pro Lys Trp Glu Lys Leu Thr Asp Ala Thr Val Trp
 355      360      365
Lys Asp Ala Ala Thr Gln Ile Phe Phe Ser Leu Ser Ala Ala Trp Gly
 370      375      380
Gly Leu Ile Thr Leu Ser Ser Tyr Asn Lys Phe His Asn Asn Cys Tyr
 385      390      395      400
Arg Asp Thr Leu Ile Val Thr Cys Thr Asn Ser Ala Thr Ser Ile Phe
 405      410      415
Ala Gly Phe Val Ile Phe Ser Val Ile Gly Phe Met Ala Asn Glu Arg
 420      425      430
Lys

```

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:

- 65 -

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

CCNAARGARA TGAAYAARCC NCC

23

- (2) INFORMATION FOR SEQ ID NO:38:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 23 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

GCNGTGAAGT ACACCACTTT NCC

23

- (2) INFORMATION FOR SEQ ID NO:39:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 23 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

CCNAARGARA TGAAYAARCC NCC

23

- (2) INFORMATION FOR SEQ ID NO:40:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

GGCYTCNGGG TAARCCACRA ANG

24

- (2) INFORMATION FOR SEQ ID NO:41:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 30 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

CGGTTCAATC TGTGTCCGC ATCAGACATG

30

- (2) INFORMATION FOR SEQ ID NO:42:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

GCAGGCTCGC GCGTCCGCTG

20

- (2) INFORMATION FOR SEQ ID NO:43:
- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 33 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

- 66 -

- CCCGTATGTC GTACTCGTGA TCCTCCTCAT CCG 33
- (2) INFORMATION FOR SEQ ID NO:44:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 27 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:
- CCNCCRTGNG TDATCATNGG RAANCCC 27
- (2) INFORMATION FOR SEQ ID NO:45:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 32 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:
- CCATTCACAC TACTGGAYYA RCAYTGNGTN CC 32
- (2) INFORMATION FOR SEQ ID NO:46:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:
- CAGATTTCTT TCTCTTTATC TGCTGCATGG 30
- (2) INFORMATION FOR SEQ ID NO:47:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 29 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:
- GGRTCDATCA TRTTYTTTATA NCKYTCNCC 29
- (2) INFORMATION FOR SEQ ID NO:48:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:
- CCTGCACCAA CAGTGCCACA AGC 23
- (2) INFORMATION FOR SEQ ID NO:49:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 25 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear
(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:
- CCAAGTACCT ACGCACACAC AAGCC 25
- (2) INFORMATION FOR SEQ ID NO:50:
(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 28 base pairs

- 67 -

- (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

GGATTAATAC GGGACCATCC ACACTACT

28

- (2) INFORMATION FOR SEQ ID NO:51:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

AGCTCTGCGG GACTTGAGAG

20

- (2) INFORMATION FOR SEQ ID NO:52:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 25 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

GTACACCACT TTTCTGAAG TCTTG

25

- (2) INFORMATION FOR SEQ ID NO:53:
 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 27 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear
 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

CCTTGGTCTG CCACATTCTC AATGTTG

27

In summary, the sequences of the Sequences Listing are as follows:

SEQ ID	Type	Sequence	Corres. Clone
1	N.A.	nt 1-190	phG2-3-a
2	Protein	aa 1-63	
3	N.A.	nt 1-190	phG2-3-b
4	Protein	aa 1-63	
5	N.A.	nt 39-1254	phG2-1
6	Protein	aa 14-418	
7	N.A.	nt 39-1635	phG2-2
8	Protein	aa 14-190	
9	Protein	aa 192-545	
10	N.A.	nt 1317-1847	phG2-4-a
11	Protein	aa 440-615	

- 68 -

SEQ ID	Type	Sequence	Corres. Clone
12	N.A.	nt 1317-1847	phG2-4-b
13	Protein	aa 440-615	
14	N.A.	nt 1540-2379	phG2-7-a
15	Protein	aa 514-793	
16	N.A.	nt 1540-2379	phG2-7-b
17	Protein	aa 514-793	
18	N.A.	nt 1-2397	
19	Protein	aa 1-797	
20	N.A.	nt 1-2397	pHGT2-a
21	Protein	aa 1-797	
22	N.A.	nt 1809-2397	phG2-8-a
23	Protein	aa 604-797	
24	N.A.	nt 1809-2397	phG2-8-b
25	Protein	aa 604-797	
26	N.A.	nt 1-2397	
27	Protein	aa 1-797	
28	N.A.	nt 1-2397	pHGT2-b**
29	N.A.	nt 296-1244	phG2-9-a
30	Protein	aa 100-414	
31	N.A.	nt 296-1244	phG2-9-b
32	Protein	aa 100-414	
33	N.A.	nt 296-1244	phG2-9-c
34	Protein	aa 100-414	
35	N.A.	nt 296-1598	phG2-10
36	Protein	aa 100-532	

** SEQ ID NO:28 encodes the same protein as SEQ ID NO: 26, though with somewhat different codon usage.

The nucleic acid sequences described herein, and consequently the protein

- 69 -

sequences derived therefrom, have been carefully sequenced. However, those of ordinary skill will recognize that nucleic acid sequencing technology can be susceptible to some error. Those of ordinary skill in the relevant arts are capable of validating or correcting these sequences based on the ample description herein of methods of isolating the nucleic acid sequences in question, and such modifications that are made readily available by the present disclosure are encompassed by the present invention. Furthermore, those sequences reported herein are within the invention whether or not later clarifying studies identify sequencing errors.

While this invention has been described with an emphasis upon preferred embodiments, it will be obvious to those of ordinary skill in the art that variations in the preferred devices and methods may be used and that it is intended that the invention may be practiced otherwise than as specifically described herein. Accordingly, this invention includes all modifications encompassed within the spirit and scope of the invention as defined by the claims that follow.

- 70 -

What is claimed:

1. An recombinant nucleic acid encoding a glycine transporter having at least about 96% sequence identity with a reference sequence which is the protein sequence of SEQ ID NO:27 or with a sequence corresponding to the protein sequence of
5 SEQ ID NO:27 except that it has one or more of the following amino acid substitutions (1) Gly¹⁰² to Ser, (2) Ser¹²⁴ to Phe, (3) Ile²⁷⁹ to Asn, (4) Arg³⁹³ to Gly, (5) Lys⁴⁵⁷ to Asn, (6) Asp⁴⁶³ to Asn, (7) Cys⁶¹⁰ to Tyr, (8) Ile⁶¹¹ to Val, (9) Phe⁷³³ to Ser, (10) Ile⁷³⁵ to Val, (11) Phe²⁴⁵ to Leu, (12) Val³⁰⁵ to Leu, (13) Thr³⁶⁶ to Ile or (14) Leu⁴⁰⁰ to Pro.
- 10 2. The nucleic acid of claim 1, wherein the reference sequence is the protein sequence of SEQ ID NO:27 or with a sequence corresponding to the protein sequence of SEQ ID NO:27 except that it has one or more of the following amino acid substitutions (1) Gly¹⁰² to Ser, (2) Ser¹²⁴ to Phe, (3) Ile²⁷⁹ to Asn, (4) Arg³⁹³ to Gly, (5) Lys⁴⁵⁷ to Asn, (6) Asp⁴⁶³ to Asn, (7) Cys⁶¹⁰ to Tyr, (8) Ile⁶¹¹ to Val, (9) Phe⁷³³ to Ser or (10)
15 Ile⁷³⁵ to Val.
3. The nucleic acid of claim 1, wherein said sequence identity is at least about 97%.
- 20 4. The nucleic acid of claim 1, wherein said sequence identity is at least about 98%.
5. The nucleic acid of claim 1, wherein the nucleic acid encodes a glycine transporter having the reference sequence.
- 25 6. The nucleic acid of claim 1, comprising the nucleic acid sequence of SEQ ID NO:26 or with a sequence that varies from the nucleic acid sequence of SEQ ID NO:26 by having one or more of the following nucleotide substitutions (a) T⁶ to C, (b) G³⁰⁴ to A, (c) C³⁷¹ to T, (d) C⁵⁷¹ to T, (e) T⁸³⁶ to A, (f) A¹¹¹⁶ to G, (g) A¹¹⁷⁷ to G, (h)
30 G¹³⁷¹ to C, (i) G¹³⁸⁷ to A, (j) G¹⁸²⁹ to A, (k) A¹⁸³¹ to G, (l) G²¹⁰³ to A, (m) T²¹⁹⁸ to C, (n) A²²⁰³ to G, (o) C³⁴² to G, (p) C³⁵² to T, (q) T⁷³³ to C, (r) A⁷⁷⁷ to G, (s) G⁹¹³ to C, (t) G⁹⁵¹ to A, (u) C¹⁰⁹⁷ to T or (v) T¹¹⁹⁹ to C.

- 71 -

7. The nucleic acid of claim 1, comprising the nucleic acid sequence of SEQ ID NO:26 or with a sequence that varies from the nucleic acid sequence of SEQ ID NO:26 by having one or more of the following nucleotide substitutions (a) T⁶ to C, (b) G³⁰⁴ to A, (c) C³⁷¹ to T, (d) C⁵⁷¹ to T, (e) T⁸³⁶ to A, (f) A¹¹¹⁶ to G, (g) A¹¹⁷⁷ to G, (h) G¹³⁷¹ to C, (i) G¹³⁸⁷ to A, (j) G¹⁸²⁹ to A, (k) A¹⁸³¹ to G, (l) G²¹⁰³ to A, (m) T²¹⁹⁸ to C, or (n) A²²⁰³ to G.

8. A vector comprising the nucleic acid of claim 1 and an extrinsic promoter functionally associated therewith.

10

9. A nucleic acid encoding a glycine transporter protein having at least about 99.5% sequence identity with all or one to two contiguous portions of a reference amino acid sequence, wherein the reference sequence is SEQ ID NO:27 or an amino acid sequence corresponding to the amino acid sequence of SEQ ID NO:27 except that it has one or more of the following amino acid substitutions (1) Gly¹⁰² to Ser, (2) Ser¹²⁴ to Phe, (3) Ile²⁷⁹ to Asn, (4) Arg³⁹³ to Gly, (5) Lys⁴⁵⁷ to Asn, (6) Asp⁴⁶³ to Asn, (7) Cys⁶¹⁰ to Tyr, (8) Ile⁶¹¹ to Val, (9) Phe⁷³³ to Ser, (10) Ile⁷³⁵ to Val, (11) Phe⁷⁴⁵ to Leu, (12) Val³⁰⁵ to Leu, (13) Thr³⁶⁶ to Ile or (14) Leu⁴⁰⁰ to Pro.

20

10. The nucleic acid of claim 9, wherein the one to two contiguous sequence portions comprise at least about 600 amino acids.

11. A cell as follows:

(a) ~~transformed with a first vector according to claim 8 and comprising said~~
25 nucleic acid, or

(b) transformed with a second vector and comprising a second nucleic acid encoding a transporter protein having at least about 99.5% sequence identity with one to two contiguous portions of a reference protein sequence which is SEQ ID NO:27 or a sequence corresponding to the protein sequence of SEQ ID NO:27 except that it has one or more of the following amino acid substitutions (1) Gly¹⁰² to Ser, (2) Ser¹²⁴ to Phe, (3) Ile²⁷⁹ to Asn, (4) Arg³⁹³ to Gly, (5) Lys⁴⁵⁷ to Asn, (6) Asp⁴⁶³ to Asn, (7) Cys⁶¹⁰ to Tyr, (8) Ile⁶¹¹ to Val, (9) Phe⁷³³ to Ser, (10) Ile⁷³⁵ to Val, (11) Phe⁷⁴⁵ to Leu, (12) Val³⁰⁵ to
30

- 72 -

Leu, (13) Thr³⁶⁶ to Ile or (14) Leu⁴⁰⁰ to Pro, wherein the encoded protein has glycine transporter activity.

12. A method of producing a glycine transporter comprising growing the
5 cells of claim 11.

13. The method of claim 12 further comprising at least one of (a) isolating
membranes from said cells, which membranes comprise the glycine transporter or (b)
extracting a protein fraction from the cells which fraction comprises the glycine
10 transporter.

14. An glycine transporter isolated from a cell according to claim 11 and
expressed by said first or second extrinsically-derived nucleic acid

15. A method for characterizing a bioactive agent for treatment of a nervous
system disorder or condition or for identifying bioactive agents for treatment of a nervous
system disorder or condition, the method comprising (a) providing a first assay
composition comprising (i) a cell according to claim 10 or (ii) an isolated glycine
transporter protein comprising the amino acid sequence encoded by said first or second
20 extrinsically-derived nucleic acids, (b) contacting the first assay composition with the
bioactive agent or a prospective bioactive agent, and measuring the amount of glycine
transport exhibited by the assay composition.

16. The method of claim 15, further comprising comparing the amount of
25 glycine transport exhibited by the first assay composition with the amount of glycine
transport exhibited by a second such assay composition that is treated the same as the
first assay composition except that it is not contacted with the bioactive agent or
prospective bioactive agent.

17. The method of claim 15, wherein the nervous system disorder or
30 condition is one of the group consisting of (a) pain, (b) spasticity, (c) myoclonus, (d)
muscle spasm, (e) muscle hyperactivity or (f) epilepsy.

- 73 -

18. The method of claim 17, wherein the spasticity is associated with stroke, head trauma, neuronal cell death, multiple sclerosis, spinal cord injury, dystonia, Huntington's disease or amyotrophic lateral sclerosis.

5 19. A nucleic acid that hybridizes with a reference nucleic acid sequence which is SEQ ID NO:26 or with a sequence that varies from the nucleic acid sequence of SEQ ID NO:26 by having one or more of the following nucleotide substitutions (a) T⁶ to C, (b) G³⁰⁴ to A, (c) C³⁷¹ to T, (d) C⁵⁷¹ to T, (e) T⁸³⁶ to A, (f) A¹¹¹⁶ to G, (g) A¹¹⁷⁷ to G, (h) G¹³⁷¹ to C, (i) G¹³⁸⁷ to A, (j) G¹⁸²⁹ to A, (k) A¹⁸³¹ to G, (l) G²¹⁰³ to A, (m) T²¹⁹⁸
10 to C, (n) A²²⁰³ to G, (o) C³⁴² to G, (p) C³⁵² to T, (q) T⁷³³ to C, (r) A⁷⁷⁷ to G, (s) G⁹¹³ to C, (t) G⁹⁵¹ to A, (u) C¹⁰⁹⁷ to T or (v) T¹¹⁹⁹ to C, under conditions of sufficient stringency to exclude hybridizations with (a) the sequence for a rat or mouse GlyT-2 transporter or (b) the sequence for a mammalian GlyT-1 transporter.

15 20. The nucleic acid sequence of claim 19, wherein the nucleic acid is a PCR primer and the stringent conditions are PCR conditions effective to amplify a human GlyT-2 sequence but not to amplify (a) the sequence for a rat or mouse GlyT-2 transporter or (b) the sequence for a mammalian GlyT-1 transporter.

20 21. A nucleic acid of at least about eighteen nucleotides in length comprising a contiguous sequence from the coding or noncoding strand of a human GlyT-2 gene or cDNA, wherein the contiguous sequence has at least 1 nucleotide difference when compared with the rat GlyT-2 gene sequence that aligns with said contiguous sequence.

25 22. An antisense molecule comprising a contiguous sequence from a coding or non-coding strand of a human gene or cDNA for GlyT-2 which is effective when administered to a cell, tissue, organ or animal to reduce the expression of GlyT-2 in the cell or in a cell of the tissue, organ or animal, wherein the contiguous sequence has at least 1 nucleotide difference when compared with the rat GlyT-2 gene sequence that
30 aligns with said contiguous sequence.

23. The antisense molecule of claim 22, wherein the contiguous stretch is

- 74 -

included in the coding or non-coding strand of the nucleic acid sequence of SEQ ID NO:26 or of a sequence that varies from the nucleic acid sequence of SEQ ID NO:26 by having one or more of the following nucleotide substitutions (a) T⁶ to C, (b) G³⁰⁴ to A, (c) C³⁷¹ to T, (d) C⁵⁷¹ to T, (e) T⁸³⁶ to A, (f) A¹¹¹⁶ to G, (g) A¹¹⁷⁷ to G, (h) G¹³⁷¹ to C, (i) G¹³⁸⁷ to A, (j) G¹⁸²⁹ to A, (k) A¹⁸³¹ to G, (l) G²¹⁰³ to A, (m) T²¹⁹⁸ to C, (n) A²²⁰³ to G, (o) C³⁴² to G, (p) C³⁵² to T, (q) T⁷³³ to C, (r) A⁷⁷⁷ to G, (s) G⁹¹³ to C, (t) G⁹⁵¹ to A, (u) C¹⁰⁹⁷ to T or (v) T¹¹⁹⁹ to C.

24. An expression vector comprising the nucleic acid of claim 22.

10

25. A method of reducing GlyT-2 expression in a tissue or cell comprising applying to the tissue or cell (a) a nucleic acid of claim 22 in an amount effective to reduce GlyT-2 expression or (b) an expression vector for expressing said nucleic acid in said tissue or cell in an amount effective to reduce GlyT-2 expression.

15

26. A method of treating a nervous system disorder or condition comprising applying to a tissue or cell of a human patient a nervous system disorder or condition treating effective amount of a nucleic acid of claim 22 or a nervous system disorder or condition treating effective amount of an expression vector for expressing said nucleic acid in said tissue or cell.

20

27. A method for detecting whether an animal has autoimmune antibodies against a glycine transporter, the method comprising contacting an antibody preparation from the animal or a body fluid from the animal with a polypeptide antigen comprising a glycine transporter or derived from the glycine transporter.

25

1/11

Human GlyT2 RT-PCR Clones

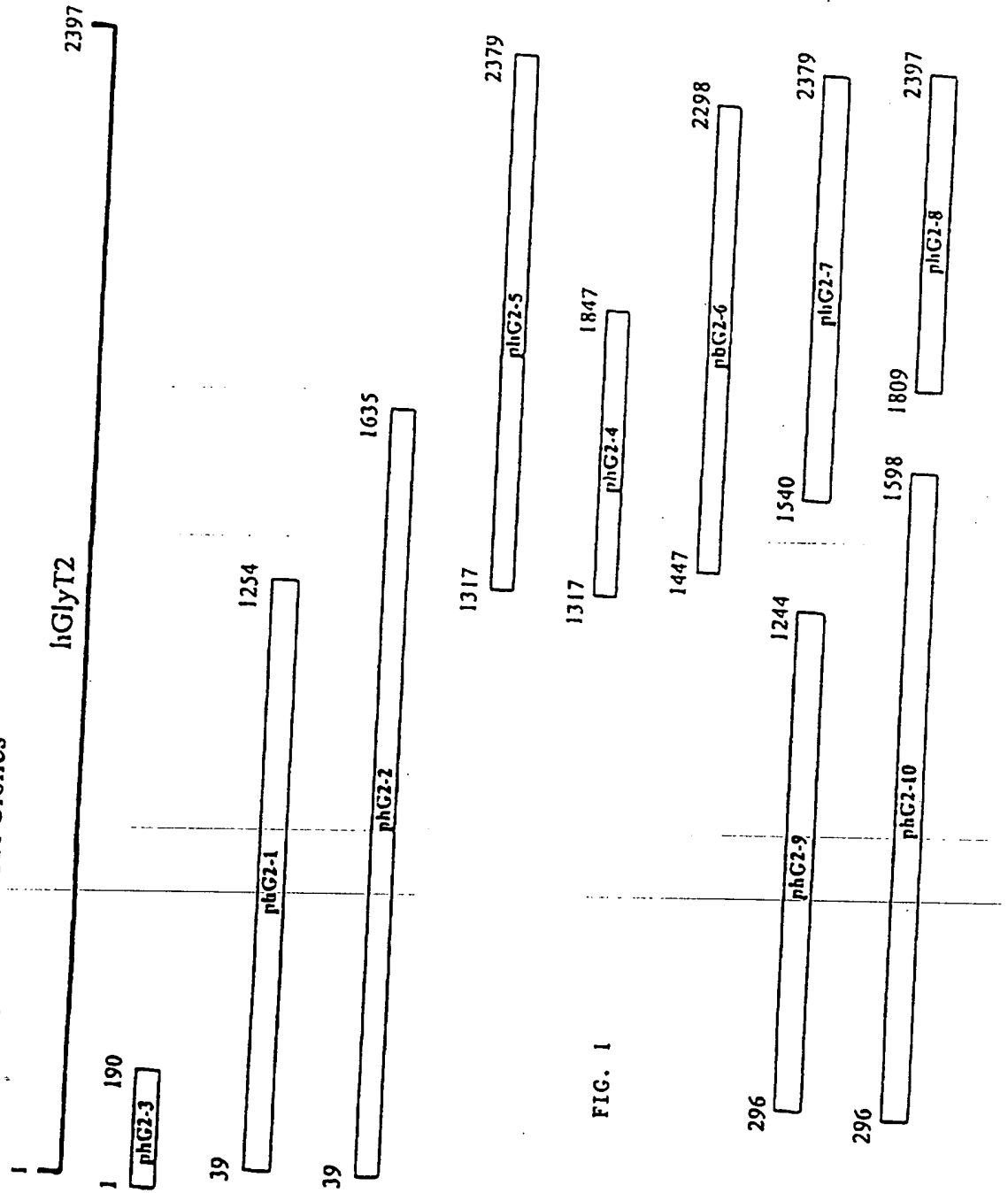


FIG. 1

Human GlyT2 cDNA

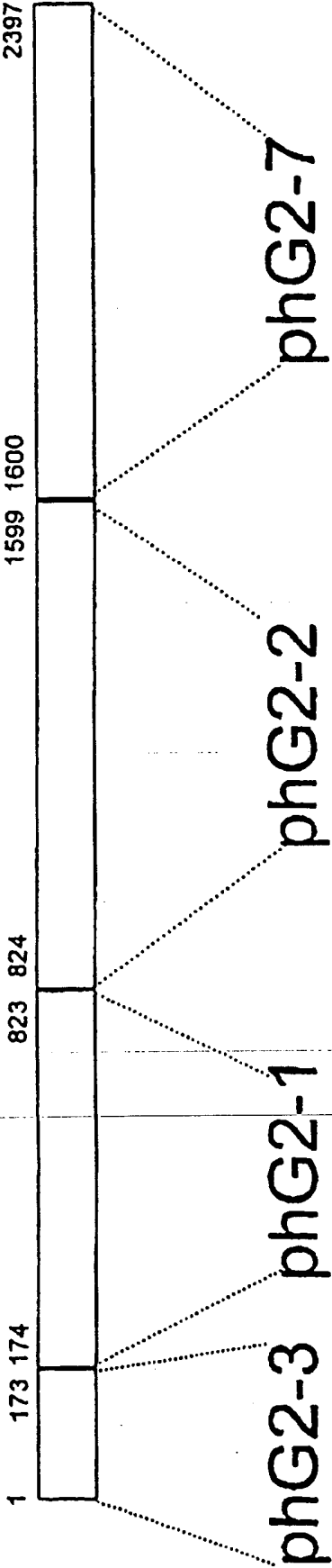


FIG. 2

3/11

Alignment of human and rat GlyT-2 cDNA Sequences

Matcht 89.0

human	10	20	30	40	50	60
	ATGGATTGCAGTGCTCCCAAGGAAATGAATAAACTGCCAGCCAACAGCCCGGAGGCGGCG					
rat	ATGGATTGCAGTGCTCCCAAGGAAATGAATAAACCACCAACCAACATCTTGGA---GGCA					
	210	220	230	240	250	260
human	70	80	90	100	110	120
	GCGGCGCAGGGCCACCCGGATGGCCCATGCGCTCCCAGGACGAGCCCGGAGCAGGAGCTT					
rat	ACGGTGCCGGGCCACCCGGATAGCCCTCGAGCACCTAGGACCAGCCCTGAGCAGGATCTT					
	270	280	290	300	310	320
human	130	140	150	160	170	
	CCCGCGGCTGCCGCC-CCGCCGC-----CGCCACGTGTGCCCAGGTCCGCTTCCACC					
rat	CCTGCGGCAGCCCCGCGGCGGCTGTCCAGCCGCCACGTGTGCCCAGGTCCGCTTCCACC					
	330	340	350	360	370	380
human	180	190	200	210	220	230
	GGCGCCCAAACCTTCCAGTCAGCGGACGCGCGAGCCTGCGAGGCTGAGCGGCCAGGAGTG					
rat	GGCGCCCAAACCTTCCAGTCTGCGGATGCGAGAGCCTGTGAGGCACAGCGGCCTGGAGTA					
	390	400	410	420	430	440
human	240	250	260	270	280	290
	GGGTCTTGCAAACCTCAGTAGCCCGGGGCGCAGGCGGCCTCTGCAGCTCTGCGGGACTTG					
rat	GGGTTTGTAAACCTTAGCAGCCCCCAGGCACAAGCGACCTCTGCGGGCCCTCCGGGACTTA					
	450	460	470	480	490	500
human	300	310	320	330	340	350
	AGAGAGGCGCAAAGCGCGCAGGCCTCGCCCCCTCCCGGGAGCTCCGGGCCCCGCAACGCG					
rat	AGCGAAGGGCACAGCGCACAGGCCAATCCCCCTCCCGGGCCGCTGGGGCTGGCAACGCT					
	510	520	530	540	550	560
human	360	370	380	390	400	410
	CTGCACTGTAAGATCCCTTCTCTGCGAGGCCCGGAGGGGGATGCGAACGTGAGTGTTGGG					
rat	TTACACTGCAAGATTCCAGCTCTGCGTGCGCCCGGAGGAGGACGAGAACGTGAGTGTTGGC					
	570	580	590	600	610	620
human	420	430	440	450	460	470
	AAGGGCACCCCTGGAGCGGAACAATACCCCTGTTGTGGGCTGGGTGAACATGAGCCAGAGC					
rat	AAGGGCACGCTGGAGCACAACAATACCCACCCGTTGGGCTGGGTGAATATGAGCCAGAGC					
	630	640	650	660	670	680

FIG. 3
(1 of 5)

4/11

FIG. 3
(2 of 5)

	1020	1030	1040	1050	1060	1070
human	AAAATACAGATCAAGAACTCGACTTTCTGTCATGACCGCTTATCCCAACGTGACAATGGTT					
rat	AAGATACAGATCAAGAACTCTACTTTCTGTCATGACTGCCTATCCGAACTTGACCATGGTT					
	1230	1240	1250	1260	1270	1280
human	1080	1090	1100	1110	1120	1130
	AATTTACCAGCCAGGCCAATAAGACATTGTGTCAGTGGAAGTGAAGAGTACTTCAAGTAC					
rat	AACTTACCAGCCAGGCCAATAAGACATTGTGTCAGCGGGAGTGAAGAGTACTTCAAGTAC					
	1290	1300	1310	1320	1330	1340
human	1140	1150	1160	1170	1180	1190
	TTTGTGCTGAAGATTTCTGCAGGGATTGAATATCCTGGCGAGATCAGGTGGCCACTAGCT					
rat	TTTGTGCTGAAGATTTCTGCAGGGATTGAATATCCTGGTGAGATCAGGTGGCCCTTGCCG					
	1350	1360	1370	1380	1390	1400
human	1200	1210	1220	1230	1240	1250
	CTCTGCCTCTTCTGGCTTGGGTCATTGTGTATGCATCGTTGGCTAAAGGAATCAAGACT					
rat	TTCTGCCTTTTCTGGCTGGGTGATTGTATATGCATCGCTGGCAAAGGAATTAAGACA					
	1410	1420	1430	1440	1450	1460
human	1260	1270	1280	1290	1300	1310
	TCAGGAAAAGTGGTGTACTTCACGGCCACGTTCCCGTATGTCGTACTCGTGATCCTCCTC					
rat	TCAGGAAAAGTGGTGTACTTCACAGCCACCTTCCCTTATGTCGTCTGTCATCCTCCTC					
	1470	1480	1490	1500	1510	1520
human	1320	1330	1340	1350	1360	1370
	ATCCGAGGAGTCACCTGCCTGGAGCTGGAGCTGGGATCTGGTACTTCATCACACCCAAG					
rat	ATTCGAGGGGTACACCTGCCTGGAGCTGGAGCCGGTATCTGGTACTTCATCACACCTAAG					
	1530	1540	1550	1560	1570	1580
human	1380	1390	1400	1410	1420	1430
	TGGGAGAAAACCTACGGATGCCACGGTGTGGAAGATGCTGCCACTCAGATTTTCTTCTCT					
rat	TGGGAGAAAACCTACGGATGCCACGGTGTGGAAGGATGCAGCCACTCAGATTTTCTTCTCC					
	1590	1600	1610	1620	1630	1640
human	1440	1450	1460	1470	1480	1490
	TTATCTGCTGCATGGGGAGGCTGATCACTCTCTTCTTACAACAAATCCACAACAAC					
rat	CTGTCTGCGGCTGGGGAGGGCTCATCACTCTTTCTTCTTACAACAAATCCATAACAAC					
	1650	1660	1670	1680	1690	1700
human	1500	1510	1520	1530	1540	1550
	TGCTACAGGGACACTCTAATTGTACCTGCACCAACAGTGCCACAAGCATCTTTGCCGGC					
rat	TGCTACAGGGACAGTTAATTGTAACCTGCACCAACAGTGCCACTAGCATCTTCGCTGGG					
	1710	1720	1730	1740	1750	1760

FIG. 3
(3 of 5)

6/11

human	1560	1570	1580	1590	1600	1610
	TTCGTCATCTTCTCCGTTATCGGCTTCATGGCCAATGAACGCAAAGTCAACATTGAGAAT					
rat	TTTGTCTCTTCTCTGTCTTGGCTTCATGGCCAACGAGCGCAAAGTCAACATTGAGAAT					
	1770	1780	1790	1800	1810	1820
human	1620	1630	1640	1650	1660	1670
	GTGGCAGACCAAGGGCCAGGCATTGCATTGTGGTTTACCCGGAAGCCTTAACCAGGCTG					
rat	GTGGCTGACCAAGGGCCAGGCATTGCATTGTGGTTTACCCAGAAGCCTTAACCAGGCTG					
	1830	1840	1850	1860	1870	1880
human	1680	1690	1700	1710	1720	1730
	CCTCTCTCTCCGTTCTGGGCCATCATCTTTTCTGATGCTCCTCACTCTTGGACTTGAC					
rat	CCTCTCTCTCCATTCTGGGCCATCATCTTTTCTGATGCTTCTCAGCTTGGACTTGAC					
	1890	1900	1910	1920	1930	1940
human	1740	1750	1760	1770	1780	1790
	ACTATGTTTGCCACCATCGAGACCATAGTGACCTCCATCTCAGACGAGTTTCCCAAGTAC					
rat	ACCATGTTTGCTACCATCGAGACCATTGTGACCTCCATCTCGGATGAGTTTCCCAAGTAT					
	1950	1960	1970	1980	1990	2000
human	1800	1810	1820	1830	1840	1850
	CTACGCACACACAAGCCAGTGTCTTACTCTGGGCTGCTGCATTTGTTTCTTCATCATGGGT					
rat	CTGCGCACACACAAGCCTGTGTTTACCCTGGGCTGCTGCATCTGCTTCTTCATTATGGGC					
	2010	2020	2030	2040	2050	2060
human	1860	1870	1880	1890	1900	1910
	TTTCCAATGATCACTCAGGGTGGAAATTTACATGTTTCAGCTTGTTGGACACCTATGCTGCC					
rat	TTCCAATGATCACACAGGGTGGAAATCTACATGTTTCAGCTTGTTGGACACCTATGCTGCC					
	2070	2080	2090	2100	2110	2120
human	1920	1930	1940	1950	1960	1970
	TCCTATGCCCTTGTCATCATTGCCATTTTGGAGCTCGTGGGGATCTCTTATGTGTATGGC					
rat	TCCTATGCTCTTGTCTCATCATTGCCATATTTGAGCTTGTTGGCATCTCCTATGTGTACGGC					
	2130	2140	2150	2160	2170	2180
human	1980	1990	2000	2010	2020	2030
	TTGCAAAGATTCTGTGAAGATATAGAGATGATGATTGGATTCCAGCCTAACATCTTCTGG					
rat	TTGCAGAGGTTCTGTGAAGACATCGAGATGATGATTGGATTCCAGCCCAACATTTTCTGG					
	2190	2200	2210	2220	2230	2240
human	2040	2050	2060	2070	2080	2090
	AAAGTCTGCTGGGCATTTGTAACCCCAACCATTTTAACCTTTATCCTTTGCTTCAGCTTT					
rat	AAGGTCTGCTGGGCGTTTGTACACCGACCATTTTAACGTTTATCCTTTGCTTCAGCTTC					
	2250	2260	2270	2280	2290	2300

FIG. 3
(4 of 5)

7/11

	2100	2110	2120	2130	2140	2150
human	TACCAGTGGGAGCCCATGACCTATGGCTCTTACCGCTATCCTAACTGGTCCATGGTGCTC					
	:: ::					
rat	TATCAGTGGGAGCCCATGACCTATGGCTCCTACCGCTACCCCTAACTGGTCCATGGTGCTT					
	2310	2320	2330	2340	2350	2360
	2160	2170	2180	2190	2200	2210
human	GGATGGCTAATGCTCGCCTGTTCCGTGCTGATCCCAATTATGTTTGTGATAAAAATG					
	:::::::: :::::::::: ::::: :::::::::::::::::: :::::::::: ::::::::::					
rat	GGATGGCTGATGCTCGCCTGCTCCGTGATCTGGATCCCGATTATGTTTCGTGATAAAAATG					
	2370	2380	2390	2400	2410	2420
	2220	2230	2240	2250	2260	2270
human	CATCTGGCCCCTGGAAGATTTATTGAGAGGCTGAAGTTGGTGTGCTCGCCACAGCCGGAC					
	:::::::: ::::: ::					
rat	TATCTGGCTCCTGGGAGATTTATTGAGAGGCTGAAGTTGGTATGCTCGCCACAGCCGGAC					
	2430	2440	2450	2460	2470	2480
	2280	2290	2300	2310	2320	2330
human	TGGGGCCCATCTTAGCTCAACACCGCGGGGAGCGTTACAAGAATATGATCGACCCCTTG					
	::					
rat	TGGGGCCCATCTTAGCTCAGCACCGCGGGGAACGCTACAAGAATATGATCGACCCCTTG					
	2490	2500	2510	2520	2530	2540
	2340	2350	2360	2370	2380	2390
human	GGAACCTCTTCCTTGGGACTCAAAGCTGCCAGTGAAGGATTTGGAAGTGGGCACTCAGTGC					
	:::::::: ::::: ::					
rat	GGAACCTCGTCCCTGGGACTCAAGCTGCCAGTGAAGGATTTGGAAGTGGGCACTCAGTGC					
	2550	2560	2570	2580	2590	2600
human	TAGTCC					
	::::::					
rat	TAGTCC					
	2610					

FIG. 3
(5 of 5)

	10	20	30	40	50
human	MDCSAPKEMNKLFPANSPEAAAAQGHDPGCPARTSPEQELPAAA---APPPRVPRVSAST				
rat	MDCSAPKEMNKPPTNILEATVP-GHRDSPRAPRTSPEQDLPAAPAAAVQPPRVPRVSAST				
	10	20	30	40	50
human	60	70	80	90	100
rat	60	70	80	90	100
human	GAQTFQSADARACEAERPGVGSKLSSPRAQAASAALRDLREAQSAQASPPPGSSGPGNA				
rat	GAQTFQSADARACZAQRPGVGFKLSSPQAQATSALRDLSEGHSAQANPPSGAAGAGNA				
	60	70	80	90	100
human	120	130	140	150	160
rat	120	130	140	150	160
human	LHCKIPSLRGPEGDANVSVGKGTLERNNTPVVGWVNMSQSTVVLGTDGITSVLPGSVATV				
rat	LHCKIPALRGPEEDENVSVAKGTLEIHNTPPVVGWVNMSQSTVVLGTDGIASVLPGSVATT				
	120	130	140	150	160
human	180	190	200	210	220
rat	180	190	200	210	220
human	ATQEDEQGDNKARGNWSSKLDFILSMVGAVGLGNVWRFPYLAFFQNGGGAFLIPYLMML				
rat	TIPQEDEQGDNKARGNWSSKLDFILSMVGAVGLGNVWRFPYLAFFQNGGGAFLIPYLMML				
	180	190	200	210	220
human	240	250	260	270	280
rat	240	250	260	270	280
human	ALAGLPIFFLEVSLGQFASQGPVSVWKAIPALQCGIAMLIISVLIATYNNVYIICYTLFY				
rat	ALAGLPIFFLEVSLGQFASQGPVSVWKAIPALQCGIAMLIISVLIATYNNVYIICYTLFY				
	240	250	260	270	280
human	300	310	320	330	340
rat	300	310	320	330	340
human	LFASFVSVLPWGSCNNPNWNTPECKDKTKLLLDSCVIGDHPKIQIKNSTFCMTAYPNVIMV				
rat	LFASFVSVLPWGSCNNPNWNTPECKDKTKLLLDSCVIGDHPKIQIKNSTFCMTAYPNLIMV				
	300	310	320	330	340
human	360	370	380	390	400
rat	360	370	380	390	400
human	NFTSQANKTFVSGSEEFKYFVLKISAGIEYPGEIRWPLALCLFLAWVIVYASLAKGIKT				
rat	NFTSQANKTFVSGSEEFKYFVLKISAGIEYPGEIRWPLPLPCLFLAWVIVYASLAKGIKT				
	360	370	380	390	400
human	420	430	440	450	460
rat	420	430	440	450	460
human	SGKVVYFTATFPYVVLVILLIRGVTLPGAGAGIWFYFITPKWEKLTIDATVWKAATQIFFS				
rat	SGKVVYFTATFPYVVLVILLIRGVTLPGAGAGIWFYFITPKWEKLTIDATVWKAATQIFFS				
	420	430	440	450	460

FIG. 4
(1 of 2)

9/11

	480	490	500	510	520	530
human	LSAAWGGLITLSSYNKFHNNCYRDTLIVTCTNSATSIFACFVIFSVIGFMANERKVNIN					
	::					
rat	LSAAWGGLITLSSYNKFHNNCYRDTLIVTCTNSATSIFAGFVIFSVIGFMANERKVNIN					
	480	490	500	510	520	530
	540	550	560	570	580	590
human	VADQGPPIAFVVYPEALTRLPLSPFWAIIFFLMLLTGLDITMFATITIVTSISDEFPKY					
	::					
rat	VADQGPPIAFVVYPEALTRLPLSPFWAIIFFLMLLTGLDITMFATITIVTSISDEFPKY					
	540	550	560	570	580	590
	600	610	620	630	640	650
human	LRTHKPVFTLGCCICFFIMGFPMITQGGIYMFQLVDTYAASYALVIIAIFELVGISYVYG					
	::					
rat	LRTHKPVFTLGCCICFFIMGFPMITQGGIYMFQLVDTYAASYALVIIAIFELVGISYVYG					
	600	610	620	630	640	650
	660	670	680	690	700	710
human	LQRFCEIEMMIGFQPNIFWKVCWAFVTPITLTFILCFSFYQWEPMTYGSYRYPNWSMVL					
	::					
rat	LQRFCEIEMMIGFQPNIFWKVCWAFVTPITLTFILCFSFYQWEPMTYGSYRYPNWSMVL					
	660	670	680	690	700	710
	720	730	740	750	760	770
human	GWLMLACSVIWIPIIMFVIKMYLAPGRFIERLKLVCSPQPDWGPFLAQHRGERYKNMIDPL					
	::					
rat	GWLMLACSVIWIPIIMFVIKMYLAPGRFIERLKLVCSPQPDWGPFLAQHRGERYKNMIDPL					
	720	730	740	750	760	770
	780	790				
human	GTSSLGLKLPVKOLELGTQC					
	::::::::::::::::::::::::::::::::					
rat	GTSSLGLKLPVKOLELGTQC					
	780	790				

FIG. 4
(2 of 2)

10/11

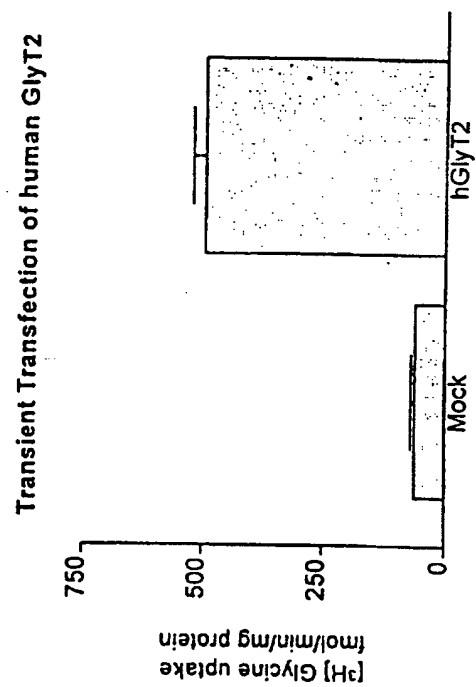


FIG. 5

11/11

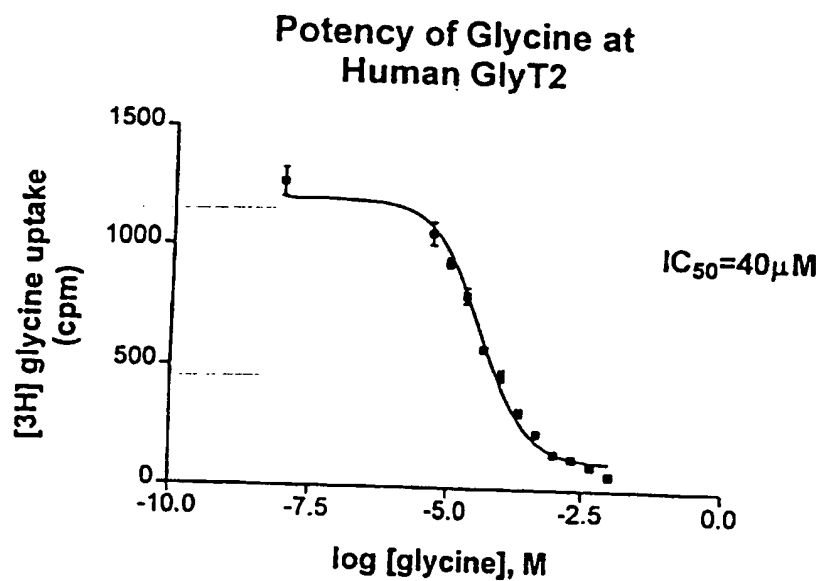


FIG. 6

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/14637

A. CLASSIFICATION OF SUBJECT MATTER

IPC(6) : C12N 15/12, 15/85; C07K 14/435; C07H 21/04

US CL : 536/23.1, 23.5, 24.33; 435/69.1, 325, 320.1; 530/350

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 536/23.1, 23.5, 24.33; 435/69.1, 325, 320.1; 530/350

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

APS, MEDLINE, BIOSIS, IntelliGenetics

search terms: glycine transporter#, GlyT2, Gly T2

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	LIU ET AL. CLONING AND EXPRESSION OF A SPINAL CORD- AND BRAIN-	21
---	SPECIFIC GLYCINE TRANSPORTER WITH NOVEL STRUCTURAL FEATURES.	---
A	THE JOURNAL OF BIOLOGICAL CHEMISTRY. 25 OCTOBER 1993. VOL. 268. NO. 30. PAGES 22802-22808.	1-14, 19-20
A	KIM ET AL. CLONING OF THE HUMAN GLYCINE TRANSPORTER TYPE 1: MOLECULAR AND PHARMACOLOGICAL CHARACTERIZATION OF NOVEL ISOFORM VARIANTS AND CHROMOSOMAL LOCALIZATION OF THE GENE IN THE HUMAN AND MOUSE GENOMES. MOLECULAR PHARMACOLOGY. 1994. VOL. 45. PAGES 608-617.	1-14, 19-21

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:	*T	later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
A document defining the general state of the art which is not considered to be of particular relevance	*X*	document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
E earlier document published on or after the international filing date	*Y*	document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
L document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	*A*	document member of the same patent family
O document referring to an oral disclosure, use, exhibition or other means		
P document published prior to the international filing date but later than the priority date claimed		

Date of the actual completion of the international search	Date of mailing of the international search report
30 SEPTEMBER 1997	29 OCT 1997

Name and mailing address of the ISA/US
Commissioner of Patents and Trademarks
Box PCT
Washington, D.C. 20231
Facsimile No. (703) 305-3230

Authorized officer

ROBERT C. HAYES, PH.D.

Telephone No. (703) 308-0196

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/14637

Box I Observations where certain claims were found unsearchable (Continuation of Item 1 of first sheet)

This international report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of Item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

Please See Extra Sheet.

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4. ☒ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
1-14 and 19-21

Remark on Protest

- ☐ The additional search fees were accompanied by the applicant's protest.
☐ No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/US97/14637

BOX II. OBSERVATIONS WHERE UNITY OF INVENTION WAS LACKING

This ISA found multiple inventions as follows:

This application contains the following inventions or groups of inventions which are not so linked as to form a single inventive concept under PCT Rule 13.1. In order for all inventions to be searched, the appropriate additional search fees must be paid.

Group I, claim(s) 1-14 and 19-21, drawn to isolated nucleic acid molecules encoding a glycine transporter protein, vectors, host cells, the protein itself, and methods of producing the glycine transporter protein.

Group II, claim(s) 15-18, drawn to methods of characterizing a bioactive agent for treatment of a nervous system disorder comprising providing a glycine transporter protein in a sample and contacting the transporter protein with the bioactive.

Group III, claim(s) 22-26, drawn to antisense DNA molecules and methods using antisense DNA molecules to reduce GlyT-2 expression in a tissue or a cell, or to treat a nervous system disorder.

Group IV, claim(s) 27, drawn to a method of detecting whether an animal has autoimmune antibodies against a glycine transporter protein.

The inventions listed as Groups I-IV do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

Group I is directed to isolated nucleic acid molecules encoding a glycine transporter protein, vectors, host cells, the protein itself, and methods of producing the glycine transporter protein, which is the first product, method of making and method of using the product. The special technical feature is the nucleic acid molecules encoding the glycine transporter protein. Groups II-IV are drawn to methods having different goals, method steps and starting materials, which do not require each other for their practice and do not share the same or a corresponding technical feature. Groups I and III are drawn to structurally different products, which do not require each other for their practice and do not share the same or a corresponding technical feature. Note that PCT Rule 13 does not provide for multiple products or methods within a single application. Since the special technical feature of the Group I invention is not present in the Group II-IV claims, and the special technical features of the Group II-IV inventions are not present in the Group I claims, unity of invention is lacking.